Species, Humans, and Transformations

Species, Humans, and Transformations

A dissertation presented

by

Enoch Allen Lambert

to

The Department of Philosophy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Philosophy

Harvard University

Cambridge, Massachusetts

August 2015

SPECIES, HUMANS, AND TRANSFORMATIONS

ABSTRACT

Do biological species have essences? The debate over this question in philosophy of biology exhibits fundamental confusion both between and within authors. In *What to Salvage from the Species Essentialism Debate*, I argue that the best way forward is to drop the question and its terms in order to make progress on two issues: how to individuate species taxa; and how to make sense of changes in explanatory frameworks across the Darwinian historical divide. I further argue that a primary motivation for anti-essentialism, biological variation, matters differently to each project.

Anti-essentialism in the philosophy of biology has inspired influential rejections of the idea that there is such a thing as human nature. In *More Bark than Bite*, I show that the arguments are significantly weaker than supposed. Moreover, none of the weighty consequences thought to follow from any genuine sense in which there is no human nature, actually do follow. The evolution-based denial of human nature has little to contribute to inquiries into the human condition, both philosophical and scientific.

Decisions about whether to undergo experiences that could change the very preference-base on which the choice is made are “transformative”. L. A. Paul argues that transformative decisions present a problem for standard decision theory when approached in a way that leans on evaluations of the experiential consequences of the choices. Her solution proposes that we approach such decisions by asking ourselves how much we value the kind of discovery involved in transformative experience. In *Shifting*

Attention on Transformative Choice, I present two problems for her solution and offer an alternative. Transformative decisions may be rationally approached by asking ourselves how much we judge the activities on the other side of transformation to be worthwhile. This proposal helps make better sense of our relationship to experiential consequences of transformation, which is more flexible than Paul acknowledges.

CONTENTS

ACKNOWLEDGMENTS

It took a lot from a lot of people for me to get to grad school and then make it through eight years of it. It is a pleasure to express my gratitude. It will be inadequate, but here goes…

Let me begin with my advisor and committee. Thanks most of all for their generosity of time and their efforts in helping me. Each has influenced my philosophical development in critical ways and I will apply lessons learned from them throughout whatever philosophical career I am fortunate to have.

Thanks to my advisor, Ned Hall, for taking me on so late in the game. Thanks to him for his confidence-inspiring encouragement; guidance in helping me find and develop my own voice in these matters; and for excellent and enlightening suggestions in the formulation and expression of my ideas. Ned is a wonderful person to have philosophical discussions with and I count myself deeply lucky that he insisted on having so many with me. I owe him much thanks as well for guiding me through several bumps along the way, both significant and small.

Thanks to Bernhard Nickel for challenging my whole perspective on the issues I address in such a constructive way. He made crucial objections and recommendations that significantly altered my approach on each paper. Whether it shows, I believe my thinking on these matters has considerably improved as a result of wrestling with the incisive feedback he gave. Bernhard also offered great advice on numerous occasions that helped me see the big picture more clearly.

Thanks to Susanna Siegel for taking on a project quite removed from her normal focus and for providing critical support and encouragement for getting over the hump,

especially with the last two papers. Her advocacy and pursuit of "philosophical" or "highly theoretical" psychology in several seminars over the years without regard for disciplinary boundaries has been really influential on me and my work (current and future), and I am grateful to her for that.

Thanks to Peter Godfrey-Smith for inspiring my interest in the topics of the first two papers and for pushing me to think big and to follow my interests. I have arrived at conclusions that oppose some of his that kick-started the initial project. But I owe much of that process to him and his singular example in my early graduate years of openness to contrary evidence and willingness to be swayed. His classes, early guidance, and approach to philosophy remain deeply influential on me and I am grateful for the opportunity to have studied with him during his time at Harvard.

Many other professors at Harvard have contributed to my development or supported me in important ways. Foremost on the list is Sean Kelly, whom I wish to think of as an honorary committee member. One conversation with him, which he may not remember, made the scales fall from my eyes on what I refer to as "orthodoxy" in my first two papers and sent me back to the drawing board. He also provided great encouragement and insight as I turned to the topic of my third paper. Although most of our discussion over the years has not concerned the specifics of my dissertation, they have been greatly enjoyable and instructive.

While I know I have learned from and wish to thank each member of the Harvard philosophy department, I must especially mention Selim Berker, Matt Boyle, Warren Goldfarb, Richard Moran, Mark Richard, Susanna Rinard, Tommie Shelby, and Alison

Simmons. Each has provided important guidance through courses (sometimes with me as a TF), dissertation workshops, or discussion.

Philosophy graduate school can be a treacherous road to follow and I want to give a shout out to all my fellow travelers along the way. I especially want to thank Diana Acosta, Olivia Bailey, Jonny Beale, James Bondarchuk, Austin Booth, Lauren Davidson, Byron Davies, Jeremy Fix, Marc Gasser, Paul Julian, Doug Kremm, Alison Kuklok, Adrian Kwek, Dave Langlois, Celine Leboeuf, Arnon Levy, Patricia Marechal, Emily McWilliams, Elizabeth Miller, Oded Naaman, Alex Prescott-Couch, Shantia Rahimian, Wendy Salkin, Kranti Saran, Zeynep Soysal, Aleksy Tarasenko-Struc, and Kate Vredenburgh. Their philosophical companionship and friendship helped me navigate much more smoothly and sanely than I would have otherwise.

Thanks to all my students who have helped me learn how to teach and how to do philosophy.

Thanks to Ruth Kolodney, Emily Ware, Vivian McLemore, Jasone Pannone, Eric Johnson-Debaufre, Nyasha Borde, and Veronica Bailey for indispensable help and guidance in jumping through the hoops necessary to access the department's and Harvard's resources. I enjoyed each of their company.

The Harvard Philosophy Department, Graduate School of Arts and Sciences, and the Mind, Brain, and Behavior initiative provided funding and excellent opportunities for teaching.

Fortunately for me, the Harvard psychology department is very welcoming to philosophy students. From William James Hall I want to especially thank Felix Warneken for a wonderful seminar and for allowing me to subsequently join his lab for a

semester. He also sponsored me for my secondary field in MBB. I was fortunate enough to TF for Josh Greene one semester and learned much from him, about teaching and about philosophy. Thanks to Elizabeth Spelke for a great seminar and for encouraging my attempts at designing some experiments concerning the development of Theory of Mind.

The psych department also has some amazing grad students whom I was lucky enough to work with, learn from, and befriend. I'm grateful to Katie McAuliffe, Amy Skerry, and Lindsey Powell for a fun reading group on animal cognition and the collaborative work it lead to. Thanks to Samuel Mehr for all his help in heading up the MBB steering committee and to all those who worked with us on it. Thanks to Regan Bernhard for helping make my TF'ing experience in psychology so fun. Thanks to Anna Leshinskaya and Alek Chakroff for their friendship and for being such fine exemplars of how to do good work while having a blast in grad school.

Susan Zawalich runs an amazing organization for graduate students through Dudley House and I am grateful to her for the care and effort she puts in, as well as for the opportunity to be a Dudley Fellow. It expanded my graduate community immensely and I made some lifelong friends.

Rosa Cao, Andree-Anne Cormier, Philip Johnston, Justin Junge, Peter Koellner, Jordane Lane, Naseem Surhio, Keith Stone, Julianne VanWagenen—Thank You. I couldn't have made it without you. Your friendships are the best things I got at Harvard.

By the Grace of the universe, I was taken in by the Kilombo Novo family which trained me in Capoeira Angola and gave me an education, unlike any other, in how to be a human. I have interacted with so many different beautiful souls through KN over the

years, it is impossible for me to list them all here. But I must mention a few: most of all Courtney Grey, better known as Zumbi, whose leadership, vision for a better world, sense of urgency for those in need, and pedagogical abilities are wonders to behold. What's most inspiring is that none of his abilities or admiration from others ultimately matters to him, only insofar as they are part of a thriving community that lifts people up and heals them. Lena Rojas, Nikoi Coley, Royal Nunes, Jake Giberson, Winston Cox, Irvienne Goldson, Celina Barrios-Millner, Amon Milner, An Duong, Jenna Parafinczuk, Anjana Sharma, Safaa Usmani, Sabrina Acloque, Kaitlyn Jolly, Armani White, Michal Shapiro, Aslin Perez, Aidid Brayboy, Xango are among those that have especially touched and enriched my life. From all of them, and from KN as a whole, I've learned more about "human nature" than anything I could in the academy. My recent work on transformative experience also owes much to them for inspiration and for my basic approach to it. Thanks also to Monalisa Gharavi for introducing me to KN and for her friendship.

I'm grateful to the Homer gang for the best damn place to live over the last six years. Thanks to Frisbay for welcoming me into the fold. Thanks to Brook Boardman for being such a great friend and tolerant roommate. He is a shining example of loyalty and steadfastness in friendship, and I'll miss him. Thanks to Colleen Sullivan, Carlie, Matthew Gibb, Amelia Evans, Meghan Frederico, Kathleen Hennessy, Dave Rod, Jennifer Vannest for keeping Homer so funky and fresh. I also count Courtney Cook, friend extraordinaire, as part of the Homer gang, since our friendship began there. Thanks to all of them for support throughout the years. Colleen's right: there's no place like Home(r).

Ithaca was a wonderful place to grow up and I continue to be sustained by the friendships and bonds I formed there. Special thanks to Micaela Kimball, Adam and Kris, Margaret Dyer, Jeff and Anita Weiss, Micah and Heather Howard, John and Marion Rantanen, Brandon Boyd, Jenny and Scott Sanders, Dave Majeroni, Anita Vaughn, Annie Grant, Alex Avenius, Ben Boruchowitz, Paul and Karen Cardon, Joe and Jackie Cassaniti, Julie Cassaniti. Nathaniel Weiss has been an especially good friend, stepping up beyond measure in my deepest time of grief.

The Birch Creek Service Ranch has been one of the most meaningful things in my life. Without its summer healing and renewal I don't know if I would have come back to grad school after my first year. I'm grateful to Adam Bateman and Eric Peterson for their dedication to making it work and for letting me help them. There are so many connected to it and its predecessor in some way to thank: Steve and Kathy Peterson, Janna and Cory Dean, Alex Peterson, Starr Stratford, Summer Peterson and Jason Van Orman, Zina Bennion, English and Kelly Brooks, Wes Peterson, Mike Horne, John Hendrickson, Lee Holmstead, Simon, Stefano, and Elisa Martinengo, Ashley Sanders, Moe, and dozens of other campers and counselors.

Eric Peterson has been an especially understanding friend through the years and I'm grateful to him. My life would be better if I could be around him more. Thanks for help on all the "issues", Ereeec. I'm so happy for you, Shayna, and sweet little Pi.

Adam Bateman has gone from being my cool older cousin, to being my boss, to being my closest friend. He has supported me, inspired me, and journeyed with me in every aspect of my life throughout grad school like no one else. I can't thank him

enough. You have a mind and a heart like no one I know, Adam. Thank you for sharing them so fully with me.

Thanks to the Cambridge crew and especially the “Quipps”: Helene, Tony, William, and extended family for welcoming me into their lives and homes. Their support, especially these last few months, has made a stressful time so much easier and enjoyable than it otherwise would have been. Of course, I owe my connection to all of them, and so very much more, to Tallulah, whose friendship and now love has come to be the joy of my life. It is a dream come true to be able to thank her in this way. You brighten my whole world, Tallu.

Gordon Wagner was an inspiration and parental figure to me all growing up. His courageous dissertation at Cornell and subsequent life path have deeply affected me. I miss him and wish he was here to press me as to what on earth my research and thesis was going to do to help save the world. I don’t know, Gordon, but I’ll keep on trying.

Just about anyone coming from old Mormon roots will be fortunate enough to have a large extended family to thank. I can’t list them all, but I especially want to thank my closest aunts and uncles, Wayne and Pat Hall and Kim and Linda Bateman (and their families), and cousin Michael Lambert, for all their love over all the years—through the worst times and the best times.

My grandparents, Leonard and Beth James, and Theron and Verl Lambert, and “adopted” grandmother, Barbara Gainey, were pillars of my world growing up and I thank them. The three I got to be closest to, Barbara Gainey, Leonard James, and Verl Lambert, saw me through much of grad school and I’m only sorry I wasn’t fast enough to call and tell them I finally finished. Thank you for all your love and lessons.

Sara, thank you for being my sister and my friend. Thank you for all you've done for me during these last eight years. You, Kevin, Austin, and Emmy Rhae bring so much joy into my life.

My mother, Cindy Lambert, has done so much more than I can begin to say here. These last few years have been especially difficult, but we've grown closer through the hard times. I admire your continuing strength, Mom. Thank you for everything.

In one way or another my father, T. Allen Lambert, has been the most influential person on my path up to and including this thesis. I still remember when he introduced me to philosophy of science, in a very Socratic kind of way. It left a decisive impact on me. As did so much of what he taught me. While he would have preferred me doing philosophically informed science to philosophy of science, he was as supportive of my graduate studies as any one could be. How I long to discuss my current conclusions and next steps with him. And to go for another long ride together. Thank you, Dad. I miss you.

I dedicate this to my parents and to the exorcism of dissertation demons past.

INTRODUCTION

The following introduction may be more biographical than many. But I think it necessary given that what I originally envisioned was a traditional monograph-like project, and what I ended up with are three papers that I never envisioned writing, only two of which are closely connected. The narrative sheds light on the unlikely unity I identify at the end of this introduction.

My thesis project began as an attempt to build on an idea that came as a big surprise to me in graduate school. That idea was the claim that evolutionary theory shows us that there is no such thing as human nature. An even bigger surprise was the fact that this claim enjoyed a position of orthodoxy in philosophy of biology due to advocacy by some of its leading figures. The surprise was congenial, though, as I had long felt most affinity toward traditions of thought in philosophy and the human sciences that emphasized the plasticity and freedom of human being, as well as the autonomy of the sciences of mind and society. Suspicious of E.O. Wilson's sociobiology and more recent evolutionary psychology, I was enthused to learn that the seeds of their own refutation lay within their home turf. I was even more encouraged by Peter Godfrey-Smith's conjecture that perhaps a fruitful connection could be found between the evolutionary and existentialist denials of human nature. It's also congenial to my mindset to explore unlikely connections between very different lines of thought. So I set out to reinforce the evolution-based denial of human nature against common misconceptions, and to locate and exploit that hoped for connection.

To reinforce the evolutionary denial of human nature, I turned to a recent challenge to its argumentative basis. That basis is a more general species anti-

essentialism in the philosophy of biology. If no species has an essence, then the species that humans belong to does not. But Michael Devitt, in "Resurrecting Biological Essentialism," brought a vigorous challenge to the anti-essentialist orthodoxy; one that I thought needed answering. If he was right, the primary justification for using evolutionary theory to deny that humans have natures was missing. If he could be refuted, so much the better for the first aim of my project.

I've been struggling with Devitt's arguments, and the reinvigoration of the species essentialism debate, ever since. The first paper of my thesis argues for my current take on it: that it has no common target for disputants to be right or wrong about. While that conclusion rests on an argument about the best way to account for the many different positions and lines of argument that have arisen in the debate, Devitt's paper is a good microcosm of what I think plagues the debate as a whole. For while his main arguments are convincing for something, it is hard to understand why they are arguments for essentialism. And it is telling that his own explicit characterization of what essences are do not help us understand any better how his arguments count in favor of them.

Devitt's primary arguments, as I understand them, are that intraspecific variation has been exaggerated by philosophers of biology and that the many true generalizations about members of species taxa require partial explanation via their genetic causes. I agree with him on the first part, and many of his opponents in the debate agree with him on the second part (indeed, I don't know who would dispute it). But for various reasons, including the fact that essentialists like Aristotle recognized variation within species, it is hard to see how these relatively uncontroversial considerations add up to essentialism.

Taking more and more of the recent literature on species essentialism into account only reveals more ways of construing essentialism (or mere intuitive appeals to it), more ways of arguing for and against it, and more historical traditions appealed to for inspiration or motivation. This leads me to conclude that the species essentialism debate is beyond repair and should simply be abandoned. However, there are two important debates in the vicinity that deserve to be addressed on their own. The first is how species taxa are individuated. That they may be individuated is, or should not be, controversial. What is controversial is whether intrinsic traits of organisms figure in the identity conditions of some species. This is a question that is often treated as one about whether species have essences. It shouldn't be.

On the question of whether intrinsic traits of organisms figure in the identity conditions of species, one line of resistance comes from the idea that there is too much diversity within species to allow it. But profligate diversity in certain traits can coexist with uniformity in others. Whether the amount of variation within a species matters to its identity conditions is partially dependent on how much, and what kind, of uniformity also exists within a species.

The second question to be divorced from the anti-essentialism debate is how best to understand post-Darwinian changes in biological explanation. Ernst Mayr kickstarted this issue by characterizing pre-Darwinian biology as ensconced in "typological" thinking and post-Darwinian biology as engaged in "population thinking". The former kind of thinking came to be associated with essentialism, the latter with anti-essentialism. The key feature of population thinking is that in the right kinds of

populations, biological variation may be taken for granted and has a positive explanatory role to play in evolutionary processes.

The shift to population thinking is misleadingly put in terms of anti-essentialism for two reasons. First, while it is incompatible with a strongly normative understanding of biological variation typically attributed to Aristotle, it is silent on other issues falling under the "essentialism" umbrella. This makes a general attribution of anti-essentialism to population thinking too coarse-grained. Furthermore, population thinking has not been the only post-Darwinian explanatory innovation. Others have been argued to have more affinity to essentialism. Once again, if the repeating comparisons to "essentialism" are simply dropped, we may achieve a more balanced and nuanced understanding of how biological explanation has advanced in the wake of the Darwinian revolution.

We need to make progress on the issue of how organismal traits figure in the identity conditions of species taxa. We also need to make progress on understanding explanatory strategies in modern biology, and how they differ from their historical counterparts. And we need to reevaluate what roles biological uniformity and diversity play in those projects. But the use of the essentialism issue as a framework encompassing both issues has lost its usefulness. Progress is more likely without it.

While my first paper results in an "anti-essentialism" of a sort, it is not one that reinforces the evolutionary denial of human nature. If anything, it invites suspicion about whether that denial also bares the marks of confusion. That suspicion is deepened by the fact that many debates and claims in the human sciences that employ the language of "human nature" are not charitably interpreted as committed to any strong stance with respect to species essences. For instance, the question of to what extent

human psychologies are “nativist” is generally understood as one about human nature. And, as I argue, the question is largely orthogonal to the considerations that exercise the species essentialism debate.

As I acknowledge, it may be true that there is no human nature, if human nature is supposed to be equated with a biological species essence of Homo sapiens. But that is compatible with many plausible conceptions of human nature, e.g., the set of psychological traits driven to fixation by the dawn of what archaeologists call “behavioral modernity”, and that continue to be fixated in human populations today. Moreover, the fact that there is no human nature in any clear, strong essentialist sense does not imply many of the things philosophers of biology have argued for. In particular, it does not necessarily imply either ethical conclusions congruent with pro-humanist outlooks, or that human biology is irrelevant to ethical or political projects. Nor does it suggest radical restructuring of the aims and methods of human psychology and anthropology. Nor, finally, are there any interesting connections to be made with other philosophical denials of human nature, including existentialism.

Of course, by the end of this second paper, I have thoroughly undermined the original aims of my thesis project. I have done the opposite of reinforcing the evolutionary denial of human nature. And I have argued that there is no interesting connection to be found between that denial and the denials found in the writings of various existentialists. This left me with a conundrum about what to do next. I hoped to extend my results to a specific, substantive conception of human nature. This is the conception of humans as the rational animal that has been so influential in the Western tradition since at least Aristotle. What lessons might evolutionary theory have for this

tradition? Unfortunately, repeated attempts to tackle the issue failed to yield adequate progress. I remain interested in the project. But the wide diversity of responses to my initial attempts to do something with it suggest to me that the issue is even more vexed than the species essentialism one.

Fortunately, I found a different, more focused and tractable topic to address: L. A. Paul's work on transformative experience. The topic does not directly relate to the arguments and conclusions resulting from the first two papers. But it did allow me to work on issues that relate to my motivations for the original project. These motivating interests include what, if anything, distinguishes humans from other intelligent creatures; and challenges to traditional conceptions of human being as pressed by various existentialist thinkers.

While it is not an explicit topic of the paper, transformative experience, i.e., experience that changes one's sense of who one is in fundamental ways, is arguably a feature unique to human beings. So understanding transformative experience may be one path toward getting a grip on what is distinctive of human psychology. The fact that it may represent a deep challenge to our understanding of ourselves as rational beings, or at least currently influential models of what that means, is also congenial to my original motivations. Many existentialists focused on the limits of rationality for making sense of humans. So at the end of this project, I am again hopeful for the prospects of investigating themes raised by existentialist thinkers through the lens of contemporary philosophical methods.

That being said, the results of my paper on transformative experience continue the theme of denying that the connection to existentialist thought has been adequately

located. For I argue that people can and do make rational choices to undergo, or avoid, transformative experiences via evaluations of the worthwhileness of the activities that the experiences open up to them. The inaccessibility of the phenomenal character of the experiences one could have due to transformation is not an impassable obstacle to employing rational methods of decision making. But there is a result of my own solution that converges with existential themes of commitment and meaning-making. Specifically, our commitments to worthwhile pursuits, combined with our ability to learn to like or overcome dislike of novel types of experiences, render the character of novel phenomenal experience much less important to calculative decision making than some (e.g., Paul) may have thought.

If there is unity to my thesis project, then, this is how I conceive of it. The first two papers may be seen as contributions to the philosophical and scientific frameworks in which specific conceptions of human nature are evaluated. The last is a small but substantive contribution to understanding what human nature is actually like.[1]

1 Thanks to Brook Boardman for helping me meet the formatting requirements

WHAT TO SALVAGE FROM THE SPECIES ESSENTIALISM DEBATE

Introduction

Until recently in contemporary philosophy of biology, the acknowledged orthodoxy was that species have no essences.[2] The last fifteen years have seen defenses of a variety of species essentialisms along with responses from the old guard.[3] I think that species still have no essences, but not because the anti-essentialists are completely correct in their arguments nor the essentialists completely wrong in theirs. Rather, there is no common target in the debate. The best arguments of both sides are better aimed at issues that drop any reference to essences. The two most important such issues are the question of how species taxa are individuated, including what role organismal traits play; and the question of how best to understand Darwin-inspired changes in explanatory frameworks in biology. It is my aim here to make the case that these two issues are best kept distinct from each other, and from any question of essentialism. Nothing pertaining to the latter issue has anything to contribute to progress on the former two. Rather, attention to the essentialism issue only confuses the genuine issues. The idea that there is a clear thesis of species essentialism worth debating is an illusion that engenders confusion.

It's quite helpful to my case that the origin of the anti-essentialist orthodoxy is not unified on a particular target, but divides between the two issues that I claim should

2 Hull (1965) and Sober (1980) are the acknowledged sources in philosophy of biology, with Mayr (1975/1959) the main biological inspiration for Sober.

3 Neo-Essentialists: Boulter (2012), Boyd (1999), Devitt (2008), Dumsday (2010), Griffiths (1999), LaPorte (2004), Okasha (2002), Walsh (2006), Wilson (1999); Counter-Revolutionaries: Ereshefsky (2010), Lewens (2012), Nanay (2011) Pedroso (2012) Rippiel (2010)

be kept distinct. The two acknowledged classics of species anti-essentialism in philosophy of biology are Hull's "The Effect of Essentialism on Taxonomy: Two Thousand Years of Stasis" (1965) and Sober's "Evolution, Population Thinking, and Essentialism" (1980), which develops Ernst Mayr's (1959/1975) distinction between "typological" and "population" thinking in a particular direction. Hull's work exclusively focuses on questions of taxa individuation and the definition of the species category. Mayr's distinction plays no role in his arguments, and explanatory issues in biology make no contribution. Sober, on the other hand, understands what is worth making sense of in Mayr's distinction in terms of a profound reorientation in how explanation works in biology. I will have more to say about the differences between Hull and Sober later on. Both papers broach broad, intricate issues. Unfortunately, the two authors' employment of the "essentialism" label traded clarity for rhetorical punch. And this lack of clarity has persisted. As my discussion of the current species essentialism literature shows, the blending of issues over taxa individuation with issues over changes in biological explanation under the banner of species essentialism is a continuing stumbling block to achieving clarity and progress on what are distinct issues. Simply removing the essentialism issue from the equation can help clear up the confusion.

Confused Debates

When a debate is confused, progress by the usual means is not an option. A confused debate won't be decided by additional evidence, a new counterexample, discovery of a position's undesirable consequences. There is some other kind of obstacle, some kind of "intellectual befuddlement", standing in the way. It is hard to say exactly

what this means though, as there is more than one way to be confused. Perhaps it is like Tolstoy's insight into unhappy families—each confused debate is confused in its own way.[4] In the case of the species essentialism debate, the confusion results from implicitly assuming that there is a clear and distinct concept of essence that everyone is arguing over, when no such common target exists.

I plan to show that there is no such shared understanding of what "species essentialism" is by examining some of the more influential contributions to the debate. The general lines of evidence leading me to this conclusion include: argumentative reliance on casual uses of the concept that are discordant with applications of explicit characterizations of it; disagreement over whether an agreed upon empirical fact or phenomenon counts as an instance of an essence; wide divergence in what authors take to be the core issues; failure to acknowledge or address such divergences; a striking correlation between number of new entries into the debate and number of new issues that are taken to be central ones.[5]

If the debate is as confused as these lines of evidence suggest, the question arises of what to do about it. My proposal for overcoming the confusion is to drop the species essentialism issue entirely. In recommending this, I do not mean to suggest that there are no clear issues being mistaken for the essentialism issue. It is not simply a pseudo-problem to be dissolved. Rather, there are two important issues in the philosophy of biology, more well-defined, that get conflated with the "essentialism" issue. The two

---

4 From *Anna Karenina*: "All happy families are alike; each unhappy family is unhappy in its own way".

5 Of course, new entries to a debate should be expected to say something new. But there is a difference between drawing a new distinction, giving a new argument, etc. that are directed at what others take to be the central issue; and introducing a new issue that the debate should "really" be focused on.

issues are what general principles guide species taxa individuation; and what kinds of changes and advances in biological explanation come out of the Darwinian transformation in biology. Biological variation, often taken as the primary motivation for anti-essentialism, actually matters differently to each project. So even the usual reason for anti-essentialism is not univocal in what it tells us. "Essentialism" is simultaneously too diffuse and too coarse-grained a target for the kinds of considerations that really matter to understanding species individuation and advances in biological explanation.

The rest of the paper proceeds as follows. Section three provides the promised evidence on the state of the debate. Section four considers options for how to respond to the debate and argues against more optimistic ones. Section five argues for what I call a **Salvaging** response by making some suggestions for how the individuation and explanation issues mentioned above will look like when unencumbered by "essences". Since the essentialism issue has importance outside of biological species, there may be wider consequences for my accusations of confusion beyond philosophy of biology. So in section six I offer a speculative diagnosis for the confusion that may have consequences for the "psychological essentialism" research program in psychology. I then conclude.

The Species Essentialism Literature

In this section I present evidence that the literature on species essentialism in philosophy of biology is deeply confused and confusing. From a bird's eye point of view, the most conspicuous problem is not just a lack of shared understanding of what species essentialism is supposed to be, but a diversity of approaches to answering that very

question. Many simply assume that the notion is intuitively clear and use "essence", and its cognates, with an implicit presumption of shared understanding (e.g., Griffiths 1999; Boyd 1999; Wilson 1999). Some specify functions and commitments of essentialist positions (Ereshefsky 2010a, b; Walsh 2006; Okasha 2002). Least often, there is an explicit characterization of what essences are supposed to be (Devitt 2008). Often there is no attempt to seek continuity with historical essentialisms (Devitt 2008; Boyd 1999; Wilson 1999). But when there is, it is not always the same historical tradition that is cited (compare LaPorte 2004 and Okasha 2002 with Boulter 2012 and Walsh 2006). And there is even appeal to different aspects of historical figures (compare Boulter 2012 and Walsh 2006). In short, there is far less shared understanding of the issue at stake than is noticed or acknowledged.

But there is not just unacknowledged diversity across authors. The confusion reaches down within authors as well. And this is the most telling sign that there is something deeply wrong with the debate. I attempt to show this in detail with three of the most important papers in the debate, each arguing for a general position defended by others but at odds with the other two. They are Michael Devitt's (2008) defense of "intrinsic essentialism", Samir Okasha's (2002) defense of "extrinsic essentialism", and Marc Ereshefsky's attack on all "new" essentialisms (2010a). The major theme to keep track of is the different perspectives on the relevance of individuation and explanation to the essentialism debate, and how those differences escape notice. Devitt exhibits confusion over the concept of essence by explicitly characterizing it in terms of its individuating function and then ignoring that characterization in arguing via appeal to intuitive explanatory considerations. Okasha attempts to hive off the individuating

function of essentialist positions (from any explanatory ones) and fails to recognize that what results is an implausible position that any attempt to individuate amounts to a form of essentialism. Ereshefsky explicitly, though insouciantly, recognizes the internal diversity of what he groups under his "New Essentialism" label, *as well as* its profound difference *from* "traditional essentialism". Yet he also fails to consider that his (rather brief) treatment of the diversity of issues present in his assorted group cry out for separate, in-depth investigation, distinct from the essentialism issue.

Devitt's primary target is Okasha, and Ereshefsky singles out both for special attention. Ereshefsky seems to understand best the importance of both explanatory and individuating functions to historical essentialisms (2010a, 683). This comes out in his critique of Okasha, which bears similarities to my own. What he fails to recognize, and this is representative, is the difficulty in trying to coherently combine those two strands in a stable, coherent, and defensible way. This is best shown by engaging with Devitt in detail. I turn to that task now, before addressing Okasha and Ereshefsky. Once I get through those three, I briefly address other authors to widen the inductive base on which I infer the confusion I accuse the debate of.

Devitt on Essences

Offering an explicit definition or characterization of essences does not always prevent lapses into appeals to intuitive considerations about what essences are supposed to be or do. This is true of Michael Devitt's paper defending what he calls "Intrinsic Biological Essentialism". In it he provides an explicit and precise characterization of what he takes essences to be. But none of his subsequent arguments

for species essences appeal to it. Here is his characterization, which he offers without argument or defense:

> "A property P is an essential property of being an F iff anything is an F partly in virtue of having P. A property P is the essence of being an F iff anything is an F in virtue of having P. The essence of being F is the sum of its essential properties" (345).

On this characterization of essence, entities with essences and entities with identity conditions are co-extensive. I take individuating conditions to be ones
That may be a desideratum for a theory of essence. But more needs to be argued if essences are simply equated with identity conditions. That F's have identity conditions is or ought to be far less controversial than that F's have essences. Evidence that Devitt is simply equating the two comes from his commentary on the characterization:

> "Essences can be fully intrinsic; for example, the essence of being gold is having atomic number 79. Essences can be partly intrinsic and partly extrinsic and relational [offers pencils and pens as examples] … essences can be fully relational and extrinsic; being Australian is probably an example because it seems that anything … can have the property provided it stands in the right relation to Australia" (345-6).

Given his list of Australian things (people, buildings, expressions, etc.), Devitt appears to hold that just about everything has an essence. This makes one wonder whether Devitt is simply equating essences and conditions of individuation. If something exists, I take it to be a fairly banal point that it can be given identity conditions of some sort or another. But it is not banal to assert that it has an essence. But Devitt never argues that all things that have identity conditions also have essences.

What is also absent from his characterization of essences is any mention of explanation. Yet an appeal to the supposed explanatory power of essences is the primary move in his main argument for them. As he summarizes that argument: "…structural explanations in biology demand that kinds have essential properties" (355). But there is no argument for why his characterization of essences can provide the kinds of structural explanations biology demands.

The Shift to Explanation

"Structural explanations", for Devitt, refer to the cellular, largely genetic, causal contributions to the features of organisms that figure in true generalizations about them (347). Such generalizations include ones like African rhinos have two horns and polar bears have white fur (351). There is a point in the vicinity of Devitt's, properly construed, that is certainly true. That many non-genetic facts about species taxa have causal explanations that appeal to genetic facts about them should not be under dispute. But why does this matter to the question of essentialism?

Devitt says, with respect to generalizations about tigers having stripes, that: "… there is something intrinsic…partly in virtue of which the animal is a tiger and which causes it to be striped. That something is an essential intrinsic property" (353, emphasis added). Again, there is no question that a striped tiger has internal causes, including genetic, of its stripedness. But why is the striped entity a tiger "partly in virtue of" that fact? His explicit characterization of essences makes no mention of causes. Clearly Devitt must be relying on something else besides that characterization here.

If Devitt is relying on more than a brute, a priori intuition about what counts as a biological kind's essence, then there must be some principle that bridges his general

characterization of essence to the biological realm. He never provides one, however. He does make an argument that generalizations like "tigers have stripes" cannot be "brute facts" (352). Agreement on that point is why we turn to genetics for (partial) causes. What Devitt doesn't provide though, for instance, is an argument for why stripedness cannot, for all its non-bruteness, still itself be part of what makes something a tiger. Stripedness is also an intrinsic feature of tigers. Why could it not be part of the essence of tigerhood? If Devitt fails to provide a principled way for distinguishing between properties that are and are not essential, then he provides no reason why all the properties of an entity are not essential to it. But a theory of essences that does not distinguish non-essential properties is vacuous.

Despite not mentioning it, perhaps Devitt is assuming an essence/accident distinction, according to which many intrinsic properties of a kind are "accidental" or "non-essential" to it. If so, then perhaps having stripes is not essential to being a tiger. Something could be a tiger even if it did not have stripes. The problem with this maneuver is that one way for something to not have stripes is for it to lack the genetic causes of stripes. But, then, so far as Devitt's arguments go, something which doesn't have the genetic causes of stripes could still be a tiger and so, by his characterization of essences, genetic causes of stripes are not essential properties of tigers.

Alternatively, perhaps Devitt supposes that the genetic causes of non-genetic traits are "brute" in a way that non-genetic traits are not. Then only "brute", or perhaps better, "fundamental", features of something are essential to it. If so, we are owed a further explanation of what is "fundamental", why only genetic traits are fundamental, and why only fundamental properties can be essential.

However "brute" gets spelled out, the true generalization that tigers have the genetic traits that are causally necessary for stripes demands explanation just as much as tigers have stripes does. So, by Devitt's own explanatory demands, it does not pass the candidacy requirements for essences. Tigers having stripes-inducing-genetic-material causally depends on the transmission of genetic material across generations, which itself causally depends on the mating activities of tigers. Patterns of mating activity in tigers may in turn depend on their stripes (e.g., non-striped tigers don't mate as successfully as striped ones). If not, we know that the visible features of many species are causally necessary to the patterns of their mating activities, and so reasoning of the exact same form could be made concerning one of them. The point of the reasoning is that our drive to understand the unseen proximate causes of tiger stripes does not decisively favor them over stripes themselves for what makes something a tiger.

To sum up my argument here, Devitt appears to be assuming a model of explanation on which the provision of information about the history of causes leading up to the explanandum counts as explanation (see Lewis 1986). But such models do not assume anything about essences and Devitt has not given any reason why any one part of the causal history leading to tigers having stripes is entitled to the honorific essence of tiger.[6] It is true, as Devitt notes, that there are whole bodies of information about regularities concerning tigers that are associated with our tiger-classificatory practices. But there is no reason why any subset of that information is uniquely deserving of the title "essence".

Undermining Intuitions

6 Thanks to Ned Hall for suggesting this summary of the problem.

If the previous argument is correct, Devitt has failed to provide a principled reason for why genetic material, and not traits like being striped, count as essential properties. The argument does depend on the proposition tigers have the genetic traits that are causally necessary for stripes, though, and one might point out that, if Devitt is right, then the truth that the proposition aims at is better expressed by a theoretical identity claim.[7] Namely, the theoretical identity that tigers are (g1….gn), where g1…..gn denotes a set of genetic properties, including those causally necessary for stripes. But that there can be such theoretical identity statements is supposed to be Devitt's conclusion and the point of my arguments above are to show that he has not adequately argued for it. Rather, he has assumed such a conclusion by implicit reliance on assumptions or intuitions or something that is not part of any explicit theory of essences that he offers.

If we do appeal to intuitions, Devitt cannot claim victory. While he does not restrict essential intrinsic properties to genetics, he does restrict them to properties possessed by zygotes (or their "equivalents" for asexual species) [347]. Hence, on his view, tiger-zygotes are tigers. But, according to my intuitions, a world in which there are duplicates of tiger-zygotes, but there never have been nor ever will be developed tigers, is not a world in which there are any tigers. Even if we allow for extrinsic essential properties, e.g., genealogical ones, we can imagine a future of our own world, in which all developed tigers die out and all that remains (of relevance to tigers) are salvaged tiger-zygotes that are somehow preserved by scientists. If I consult my intuitions about

7 Nothing substantive turns on my use of a generic that may not be universally true. It is not true that all tigers have stripes. But my argument here is aimed at Devitt's framework even if there are universal truths about tigers, e.g., tigers have DNA. This sidesteps his claims that his framework can accommodate exceptions.

this possibility, I think that something "essential" is missing so that we would then live in a world without tigers, though it may be one where it is possible for tigers to return. But I only mention all of this for dialectical purposes. I do not think that intuitive judgments should decide the matter.

My arguments thus far do not rule out any intrinsic properties from being essential ones. They only show that Devitt himself has not ruled out any intrinsic properties from being essential ones, at least not in a principled way. But no one, it will be objected, will want to count all intrinsic traits of organisms of a kind as essential to the kind, for surely there can be tigers that have no stripes. My response is that such judgments may indeed be reason enough to exclude stripes from the identity conditions of tigers. But as I suggested earlier, why would anyone ever deny that tigers have identity conditions? To insist that tigers have essences because they have identity conditions is to insist on unnecessary metaphysical overlay. It is, I suggest, to be moved by an intellectual drive to connect identity and (certain kinds of) explanation to an extent that may not be warranted by the actual phenomena (more on this in the last section).

The Failure of Devitt's Response

Toward the end of his paper, Devitt briefly addresses the question of why the properties he has in mind are essential properties. He says that if all members of a species, S, have some trait, say stripedness, then they also all have the intrinsic trait(s), E, that explains stripedness.[8] Hence, all members of S have E. From here he makes the modal claim that anything that *could be* a member of S *would have* E and concludes

8 Ignoring, again, whether stripedness is actually universal. It is an easy example to use, even if it isn't strictly universal. Other examples of traits that are universal could be substituted.

that E must be an essential property of S, since "that is what it is to be an essential property" (378).

Note, however, that while Devitt's explicit characterization of essences entails that something has its essential properties wherever it exists (in possibility or 'temporality' space), it does not entail the converse. It does not entail that every property something has wherever it exists is an essential property. And that is what Devitt needs for the above argument to work. Furthermore, as he claims that the explanatory connection between stripedness and E must be law-like, any claim that universal properties of a species are essential will not distinguish between stripedness and E qua essential properties (376-7). This is a consequence of neither mentioning nor arguing for anything about explanation in his characterization of essences.

Simply amending his definition is not an option for Devitt, at least not without further argument. For instance, it may be true in our toy example that all S's are striped "in virtue of" possessing some intrinsic property E. But that is a causal sense of "in virtue of" and it is prima facie *not* the case that something being an S *in virtue of* possessing property E picks out the same kind of causal relation. We are owed more clarification of what the non-causal relation is and why some intrinsic properties and not others are candidates for being at one end of it. So granting Devitt his explanatory claims really does not get him any closer to his intuitive vision of essentialism.

There is a further hurdle for any attempt by Devitt to argue for the essentialism of certain intrinsic explanatory properties. He writes that his thesis only establishes that species essences are "at least partly" intrinsic, and indicates that he does not wish to take issue with the standard view that at least some species are historical entities

(346-7). That is, he doesn't wish to take issue with the idea that organisms sharing internal features could belong to different species taxa in virtue of spatiotemporal position and inclusion in distantly related lineages.

But if Devitt is willing to allow extrinsic or relational properties candidacy for essential traits, we also need further argument for why certain potential candidates should be ruled out. In particular, he appears not to countenance the environmental properties that are just as necessary as intrinsic properties to the development of stripes as potential essential properties. And if he is willing to countenance both relational properties and explanatory properties, why not relational properties that (help) explain what he demands explanation for? I conclude that Devitt has not adequately answered the question of why just the traits he singles out are essential properties. His identification of intrinsic essential properties of species follows neither from his explicit characterization of essences, nor from any arguments he provides.

Despite the problems I have highlighted in Devitt's position and arguments, they are not of the sort generally noted in responses to him from defenders of the anti-essentialist orthodoxy. So far these responses have tried to dispute his interpretation of biological facts or scientific practice with respect to the individuation of species (Ereshefsky 2010a and Lewens 2012).[9] This indicates implicit acceptance of his way of construing the whole essentialism issue. I think this highlights a widespread, if tacit and largely inchoate, assumption that essentialism has to fit together individuating and explanatory components *somehow*. My engagement with Devitt is meant to show just

9 Both published and most of those he mentions as personal communication. Only Peter Godfrey-Smith does he mention as challenging him on the "yes, but why essentialism?" front. I have tried to show how this line of resistance can be pushed more thoroughly than Devitt acknowledges.

how difficult actually meeting the demands of that assumption turn out to be. On the other hand, if one or the other strand of this essentialist "impulse" is disregarded, it becomes extremely dubious whether we continue to have a *single* debate, much less one where "essentialist" terminology and rhetoric contributes rather than detracts. I turn to this issue next, in my discussions of Okasha and Ereshefsky.

Okasha on Essences

The divide between Okasha (2002) and Devitt is best introduced by a quote from Devitt's paper that itself quotes Okasha. In the context from which he quotes, Okasha denies that there is a genetic property of a species that *all* members of the species possess and *all* other organisms lack. Here is Devitt, quoting and commenting on what Okasha says next:

> "This is not to deny, of course, that there are important genetic similarities between members of a single species…species taxa are distinguished by clusters of covarying [chromosomal and genetic] traits, not by shared essences" (Okasha 2002, 197)". Great! So the clusters are the essences!! [Devitt 2008, 371; original enthusiasm].

There is, or should be, something suspicious about triumphal metaphysics in the face of agreement over the relevant scientific facts. At the very least, this illustrates how differently Devitt and Okasha conceive of essences. Although he endorses much of the spirit behind the traditional rejection of essentialism, Okasha argues that species have extrinsic essences. His arguments also seem to equate essences and individuation conditions. He contends:

> "For if we cannot find a set of properties in virtue of which my pet dog Rover is a member of Canis Familiaris and not some other species, then how can we account for the facticity of assertions like "Rover is a member of Canis Familiaris"? Anyone who denied that assertion would presumably just be wrong, and we can hardly treat "being a member of Canis Familiaris" as a

> brute irreducible property that some organisms have and others do not, so how can there not be a set of essential properties which define the species?" (198-199).

To my ear, this is a perfectly good argument that species taxa have conditions of individuation. Why does Okasha present it as one for essences? He does not say. But he does go on to bare his assumption that the two sorts of things are the same while arguing for the candidacy of extrinsic properties:

> "For if the essence of a kind is simply supposed to be that set of properties which are jointly sufficient and individually necessary for being a member of a kind, then there is no particular reason why those properties should be required to be intrinsic" (202).

For Okasha, then, the essence of a kind is simply whatever it is that distinguishes members of the kind from non-members. All identity conditions are essences. There are not necessarily any special entities, essences, that provide identity conditions. But what, we should ask, is the point of equating essences and identity conditions? If essences are not something over and above identity conditions, why bring along the historical and metaphysical baggage usually associated with them? If everyone would be willing to admit that species have identity conditions of some sort or another, why argue further, for species essences? Is there any added content to calling them "essences"?

These questions become even more pressing once Okasha recognizes both the definitional and explanatory strands of essentialism and argues against the latter. He targets the Kripke-Putnam revitalization of essences as his foil. And he uses "semantic" as a label for identity conditions:

> "Hidden structure" plays both a semantic role and a causal-explanatory role, in the Kripke-Putnam story. But there is no a priori reason why the same thing should play both of these roles. It is perfectly possible that the extension

> of a kind term should be determined not by superficial characteristics but by "something else" just as Kripke and Putnam say, without it being true that that "something else" causally explains the presence of the superficial characteristics." (203).

Okasha raises a good question about the connection between kind identity and the explanation of the kind's "superficial" characteristics. And he is adamant about the answer—soon after the above quote he insists there is no reason at all (not just no a priori one) to think that the same thing need play both an individuating and explanatory role for a kind. But as Ereshefsky (2010) argues, there is certainly historical reason for thinking that if there are essences, then they do work of both kinds: every major essentialist since the original, Aristotle, has thought so. Does Okasha think that they have all been wrong about the concept of essence? If so, then presumably he thinks there is an a priori concept of essence about which we can be substantively right or wrong about? If he does think this, then we are due more argument than a simple assertion that there is no a priori reason why essences can lack explanatory power. If he does not think this, then we are due some scientific reason to posit essences.

Okasha's paper is explicitly an exercise in naturalized metaphysics—biology answers, or at least provides the material for answers to, questions about the metaphysics of species. So it is curious why he does not go a step further and question the essentialism question itself. Biologists certainly don't explicitly appeal to essences in doing their work. A naturalist approach, then, might be understood to recommend rejection of the question of essences, as either irrelevant to what science tells us about the world, or as unnecessary metaphysical gloss.

More importantly, though, the question of essences is a distraction from an important issue. The focus of Okasha's arguments for relational essentialism are on why relations can be essential properties. Otherwise, he rehearses others' reasons for why species taxa are determined by features of lineages rather than intrinsic traits of organisms. But he seems to assume that either intrinsic properties must be sufficient for species individuation or that extrinsic properties are sufficient. For he does not consider the possibility of combinations of intrinsic and extrinsic properties being necessary to do all the individuating work. He should have, because there are troubles for a purely extrinsic approach to individuating species.

The noble gases do not form the basis for life the way that carbon, hydrogen, oxygen, nitrogen, etc. do. This is because they cannot enter into the right kinds of relations with other elements. And they cannot do so because of certain intrinsic features they have. They may be related to the predominant elements of life in all kinds of ways. But they aren't related to them in the biochemical ways necessary for life. Certain intrinsic features of elements are necessary to be able to enter into biochemical relationships. Similarly for organisms on Earth and their relations of descent. Not just any relation counts as a relation of descent and not just any entity can enter in to such relations. Intrinsic properties of things matter to which ones can can be related to each other as biological "parents" and "offspring". If this is so, then relations of descent are not all that matter to individuating species.

The above issue and line of reasoning are not really considered by either Okasha or Devitt, or by others in the essentialism literature.[10] My suspicion is that this is due to

10 Devitt's arguments for intrinsic essences focuses on the explanation of "superficial" features of organisms

focus on the essentialism issue rather than the issue of individuation as such. The former issue is too much of a distraction.[11] Another place where important issues get glossed over too quickly due to concern with anything calling itself "essentialist" is in Marc Ereshefsky's paper against what he labels the "New Essentialism".

Ereshefsky Against "New" Essences

We have just seen that Okasha and Devitt fundamentally diverge over what essentialism amounts to, despite their explicit focus on just its individuating function. Other defenders of various "essentialisms" have further differences with each. In "What's Wrong With the New Biological Essentialism", Marc Ereshefsky undertakes to quash each new outcropping of essentialism, due to the fact that "they (all) concur that biological taxa have essences" (674).[12] But Ereshefsky himself offers a reason to consider a possibility he doesn't, which is the possibility that this is concurrence in label only: "As we shall see, each form of the new biological essentialism rejects one or more … features of traditional essentialism" (675 ft. 1).

As a general rule, if position X rejects core features of position Y, we either shouldn't understand X as a version of Y, or we should demand some reason for why we should understand X as a version of Y. Using the label "Y" is not usually a sufficient reason to do so. As Ereshefsky also notes that each "essentialist" position differs over both the nature of, and reasons for, essentialism, it seems like he should consider the option that distinct issues are being artificially grouped under the same heading in the

11 Although the issue also doesn't seem to be addressed in general defenses of the lineage conception of species, as in de Quieroz 1999.

12 Ereshefsky's targets include: Griffiths, Boyd, Devitt, LaPorte, Okasha, and Wilson. But not, for instance, Walsh.

guise of one determinate position. Since the positions he considers all differ over fundamentals, his own arguments are correspondingly diverse, raising the question whether he is defending a clear anti-essentialist position, or reacting to labels.

In fact, Ereshefsky's article broaches several different issues in cursory manner, each deserving of independent, and more thorough, hearing. These issues include: whether arguments usually considered by philosophers can overturn common scientific practice (676); whether intrinsic traits do, as a matter of fact, figure in common biological classificatory practices (ibid); whether or not the kind/individual distinction matters to the ontological status of the species category (678-9); how traits get explained (680); how taxon membership is determined (680-682); whether species essences can partially overlap (see 680); whether essences are supposed to do explanatory work (682-3); the role of history in determining what an essentialist position must amount to (683); and more besides.

The way I read Ereshefsky's paper, it serves as an impressive exhibition of evidence that there are persistent and important questions about how to understand species which consistently get huddled together under "the" essentialism problem. None receive adequate treatment in their own right. Each of the issues just listed makes repeated appearance in the species essentialism literature, though all are not touched on in every paper. Some papers focus on some of the issues, others on others. So there is not even agreement in the literature on what is relevant to the species essentialism issue, let alone what the most important issues are. The problem is just further compounded when we turn to additional contributions to the debate not addressed by Devitt, Okasha, Ereshefsky, or their primary interlocutors.

More Cooks in the Kitchen

Other contributors to the debate have a wider historical lens on the matter, but end up adding even further complications. Nanay (2011) understands the species essentialism issue as part of the epic historical struggle between "Realism" and "Nominalism". On the one hand, he rightly suggests that many of the considerations appealed to in the species debates are neither strong nor general enough to decide that issue. But, appealing to Ereshefsky's (2010) list of features of essentialism, he argues that there is one which evolutionary biology denies. It denies a commitment to explanation that makes appeal to biological properties Realistically understood. He argues for an interpretation of population thinking on which it makes "trope nominalistic" explanations available to biology (see 181). And this is the sense in which there can be an anti-essentialism unique to biological species that avoids the larger historical debate.

I'm not sure I fully understand Nanay's claims, or his arguments for them. They appear to rely on some kind of argument that there is a lack of appeal to certain kinds of generalizations in making biological explanations (see 185-190). They also depend on the explicitly nominalistic strand in Mayr's (159/1975) own construal of population thinking. But it has been shown that there was more than one strand of thought in Mayr's notion of population thinking, and that he was not always internally consistent (Hey 2011). Sober's (1980) pro-population thinking and anti-essentialist arguments rested on dismissing Mayr's nominalistic tendencies and bringing out what is important about populations conceived of as entities in their own right—one's about which causal-explanatory generalizations are made.

Whether Nanay or Sober are right about population thinking, Walsh (2006) would argue that both are wrong about the completeness of population thinking to our understanding of evolution. He argues that population thinking leaves out the crucial role of the organism in mediating the phenomena covered by traditional population biology (changes in gene frequency), and that this is the lesson of recent "evo-devo" thinking (see esp. 436-442). Simplifying: recent developmental biology has discovered that organismal development is surprisingly robust against many kinds of disturbances and Walsh interprets this result as both necessary to the explanation of adaptive evolution and as sufficiently similar to Aristotle's understanding of organisms to inherit his mantle. This latter part comes from the idea that Aristotle understood organisms as having a form that it was their telos to realize. This telos is their essence (see 427-430).

Walsh's sensitivity to historical matters shows that, for instance, what Ereshefsky (2001; 2010) attributes to traditional essentialism was not actually held by figures such as Aristotle, and so anti-essentialist considerations of species evolvability, stability, variation, vagueness, etc. miss their mark (interpreted as an Aristotelian mark). He does think that Aristotle demanded essences be of a simple, unified sort and so denies that certain "essentialisms", e.g. Boyd's (1999) and Devitt's (2008), really are such (see 427, and ft.4). But developmental biology, he argues, has reconfirmed Aristotle's simple essences via robust development of organisms.

Here is not the place to address all of Walsh's arguments. My own questions for him include why robustness needs to be interpreted in the heavily teleological manner that he wants to interpret it in. I also don't think that the robustness that in fact exists can do nearly as much of the explanatory work that Aristotle's telos's were meant to do.

To hold that a certain kind of population thinking is crucial but not complete, as he does, is not a sufficient basis for reintroducing a mode of thinking that was totally blind to population thinking. In addition, Dumsday (2010) takes up and defends the cause of complex essences. So the simplicity versus the complexity of essences adds yet another issue to the mix.

Finally, Boulter (2012) argues that evolutionary theory must presuppose Aristotelian essentialism. He does so on the basis of the fact that both Darwinists and Aristotelians are committed to the reality of change, and Aristotelianism, apparently, remains the best account of how there can be changes in something that remains the same thing. He justifies this move by highlighting a few of the confusions in the literature I have noted and arguing that the best way to overcome them is via return to shared "first principles" (see p. 94).

So to add to the issues identified in the debate running through Devitt, Okasha, and Ereshefsky, we also have: Realism vs. Nominalism; how to understand the explanatory power of population thinking; the "return of the organism" to biology[13]; whether said return should be interpreted teleologically; whether essences must be simple; and Aristotle vs. Parmenides.[14]

Debates presume disagreement. But they also rely on some shared understanding of the issues involved. I submit that the evidence from the species essentialism literature amounts to a compelling case that too many of the bases for shared understanding are

---

13 I owe the phrase to Nicholson (2014), without necessarily endorsing all aspects of the paper.

14 Rieppiel (2010) reveals even further issues and problems with the diverse motivations behind the variety of positions grouped under the "new essentialism" label. His paper has an anti-essentialist tone, but no unambiguous anti-essentialist conclusion. He also presumes the kind of unity in the original anti-essentialist consensus that I denied in the introduction and return to in a later section

missing for progress to be possible. There is not enough agreement on: what essences are supposed to be; what the relevant biological considerations are; how the potentially relevant biological considerations are to be interpreted. And since there is so much unresolved under each of those three headings, the absolutely crucial issue of how the biological considerations are supposed to bear on the essentialism question is a currently intractable problem. I turn next to what to do about it.

Responding to Confusion

Consider the following two ways of responding to the state of the species essentialism (SE) debate. Hopeful respondents think that the question of whether species have essences has a determinate answer and the debate is somehow decidable. The way of Despair denies this.

The way of Hope assumes that there is a clear, coherent, and non-trivial conception of essence. As a matter of minimal charity to the debate thus far, the way of Hope must also grant that this conception is not prima facie obviously applicable to biological species. It will take significant argument to decide either way. Hopefuls can regard the debate I have presented in one of two ways. They can decide that the current confusion is real but that there was a point at which the debate "went off the rails" (perhaps from the start) and we can return to that point and fix it. The other option decides that the confusion I have been documenting is only apparent or superficial. With the correct understanding of essence in hand as a hermeneutical key, we will come to see just how close everyone really is to solving the problem. Call these options Repair and Revelation, respectively.

The way of Despair takes the documented confusion as sufficient evidence that there is something wrong with the question of whether species have essences. Most likely, there is no clear, coherent, and non-trivial conception of essence for species to possess. So this way embraces anti-essentialism of a sort, but not for reasons specific to biology or the metaphysics of species. Rather, the whole debate has been a disastrous chase down a rabbit hole—disastrous due to its waste of time and talent on a pseudoproblem. Despair would agree with Repair that the debate went off the rails at some point (definitely from the start), but not because the debaters failed to correctly characterize essences. For there is nothing there to be correct or incorrect about.

I think that Hope, in both its guises, is probably wrong. But I don't think Despair is entirely correct. I think there probably isn't any coherent conception of essence worth arguing over. But I don't think there aren't any lessons to learn from the debate, or that there aren't well-defined questions in the vicinity to be answered. Instead, I think that good questions can be Salvaged and approached afresh, sans distracting concern over essentialism. The way to Salvage requires holding, against the voice of Despair, that there are some important recurring threads of argumentation in the debate that significantly track something(s). Simultaneously, it requires denying that what is tracked are the essences held out for by the voice of Hope.

The best way to argue for Salvage over Despair is to actually pick out the important threads of argument and articulate the open questions that they are best directed at. This, along with indicating that there may be better, non-essentialist arguments for addressing the questions, provides some reason against Hope. But denying Hope amounts to a negative existential (no non-trivial and coherent concepts of

essence!), which is notoriously tricky to argue for. So let me approach the path of Salvaging by saying a bit more about why I reject Hope, in either of its guises.

Against Hope

Both contemporary metaphysicians and historians of philosophy could justly complain that most participants in the debate over species essentialism do not adequately avail themselves of the resources available to get clear about the notion of essence. But only convinced advocates of one or other of the many competing positions on what essences are could honestly say that taking all those resources into account lends Hope.

The Stanford Encyclopedia of Philosophy (SEP) entry on "Essential vs. Accidental Properties" acknowledges three competing ways of drawing the distinction in the current literature, all of which are extensionally inequivalent to each other (Robertson and Atkins 2013). What gets labeled the "modal" approach is inspired most by Kripke's and Putnam's work on reference from the 70's. That is also the work often cited by various "new" essentialists (esp. Okasha 2002 and Laporte 2004). The attitude usually seems to be that new developments in the metaphysics of essence prompt re-examination of the issue with respect to species (ibid). The problem with this move is that neither Kripke (1980/1972) nor Putnam (1975, 1970) explicitly adopt a particular conception of essence from history, nor propose one of their own. Instead they employ "essence" under a presumption of shared understanding. Putnam (1975, 1970) even usually puts it in scare quotes to signal noncommittal use of the word. This has been noted by new advocates of what the entry calls "definitional" approaches to the distinction who argue that the Kripke-Putnam approach to essence does not pick out

what many historical philosophers thought essences were (Fine 1993, Charles 2003, Oderberg 2008).

The "explanatory" approach to drawing the essence/accident distinction is the third one mentioned by the SEP article, with Gorman (2005) cited as its most recent defender. This approach treats something's essence as its most fundamental explanatory properties, but without claiming that such properties are "definitional". Although the advocates of the definitional and explanatory approaches are not usually cited or appealed to in the species essentialism literature, one might think its (the literature's) diversity of emphases are reflected in the three different kinds of essentialism noted by the SEP article. If true, that doesn't lend much hope for progress in the debate. As the authors of the article say: "It is not clear whether these three characterizations should properly be thought of as competing characterizations of a single notion or instead as ways of trying to capture three related, but different … notions" (Roberston and Atkins 2013). The authors go on to say that they will focus almost entirely on the modal characterization for no other reason than historical inertia of 50 years in the analytic tradition. If this is the state of essentialism in contemporary metaphysics at large, it doesn't bode well for sorting out the species essentialism confusion. For, *at best*, the latter just reproduces the current unclarity and disagreement in the issue present in the former.

Given the current unclarity in determining what essences are, a good idea might be to turn to history for Hope in sorting out the issues. The trouble with turning to history for Hope is that the troubling controversy does not cease. For instance, in the turn from the scholastic to the modern period, Locke complained that the scholastics

had distorted Aristotle's conception of essence and was in turn criticized by Leibniz over his attempt to distinguish real and nominal essences (see Atherton 2007). But there is no doubt that Locke's microstructural real essences are shorn of many Aristotelian features of essences (e.g., natural teleology). Many of today's advocates of definitional accounts of essences advocate some kind of return to Aristotle's conception. He was, after all, the one to coin the Greek neologism that comes down to us as "essence".

When we come to Aristotle, I am not prepared to argue he had an incoherent conception of essence. But that doesn't automatically mean that Hope has found its fulfillment by returning to his conception. First, our theme continues: there remains scholarly controversy over just what Aristotle's conception is. Second, a literal appropriation of Aristotle's conception in all its detail is surely false of biological species.[15] There is evidence that Aristotle couldn't make his conception work even for the empirical biological facts that he was aware of (see Charles 2003). The route of appropriating Aristotle will have to slog through the controversy of what is sufficiently similar to him to count as anchored to his conception so as not to amount to a new invention tailored to fit what we now know about biological species. Hope has its work cut out. And a long road to travel before reaching genuinely biological issues again.

Finally, the rationalist will hold out Hope for the eventual discovery of the true contours of the a priori concept of essence. But there is no reason for the naturalistically inclined philosopher of biology trying to progress on the metaphysics of species via making sense of the Darwinian transformation of biology to await this discovery. Hence, I do not take the case against the viability of Hope to be demonstrative. But there are

15 Darwin's advance was not necessary to show this much.

considerable obstacles to its success that may not be worth pursuing, especially if a Salvaging project can be made attractive.

A Salvaging Project

A Salvaging project that rejects the essentialism issue could take a number of directions. It could reaffirm the genuine insights of the original anti-essentialist orthodoxy. It could favor some aspect of the neo-essentialist revival. My version aims at bringing questions to the fore. What debates can be refocused once concern for essentialism is gone? In the introduction, I focused on the questions of species individuation and how explanatory frameworks have changed in the history of biology. I develop that line of thought further in this section.

As a primer for how these issues can be clarified consider: insofar as purely lineage based conceptions of species gain favor or additional support from the belief that alternatives aren't viable due to essentialist entanglements, we can cast that worry aside. In the end, I do think that the distribution of organismal traits matters to the individuation of species, but the route to making that case involves raising questions of how we think about variation both within and across species. Biological variation ultimately has different roles to play in questions of explanation and questions of species individuation. Being able to distinguish them is one thing that may be usefully salvaged from the essentialism debate.

Anti-Essentialist Origins

Salvaging distinct pieces from a complicated situation with interlocking parts is difficult business. Finding a place to begin can be frustrating. In this case, history gives

us a guide. For the essentialist orthodoxy was never actually the result of one, unified line of thinking. There are two distinct strands: one tracing to David Hull's "The Effect of Essentialism on Taxonomy: Two Thousand Years of Stasis"; and one tracing to Ernst Mayr's (1975/1959) distinction between typological and population thinking, especially as developed in Eliot Sober's "Evolution, Population Thinking, and Essentialism". Each strand gave rise to different ways of rejecting essentialism. I'll refer to the traditions begun by each as the "Hull-strand" and the "Mayr/Sober-strand".[16] They could also respectively be referred to as the "individuation-strand" and the "explanatory-strand". They have important differences as well as commonalities, which I turn to next.

As his title hints, the Hull-strand is exclusively concerned with questions about the individuation of species taxa, and the nature of the species category. That initial paper highlights the difficulties that taxonomists had producing necessary and sufficient criteria for taxa membership, arguing that the continued quest to do so must be an unfortunate consequence of essentialist thinking. The quixotic nature of the quest, it was claimed, had to do with the overwhelming amount of variety to be found within species. The diversity of intrinsic traits within species was supposed to show that they could not be appealed to in order to draw species boundaries. His influential view of species as individuals wasn't developed until later, but it helped spark the idea that taxa membership could not be determined via the intrinsic traits of organisms. Instead, it is determined via relational properties, i.e., the relations of descent necessary for inclusion in a lineage.

16 It's not that Hull doesn't also cite Mayr. But Mayr is one of many biologists appealed to and for different reasons than the population thinking that Sober latches on to.

The Mayr-Sober strand is focused on explanation rather than individuation.[17] It joins the Hull-strand in taking there to be something very significant about biological variation, but differs in what the significance is. For the Mayr-Sober strand, variation has explanatory power that is unrecognized by the "Aristotelian" tradition. Properly understanding this requires significant restructuring of our biological explanatory framework. Intraspecific variation moves from explanandum in the "Aristotelian" world to explanans in the world of population biology. Consequently, statistical norms and stereotypes lose any telic or normative explanatory power and become explananda themselves.[18]

Now note that these separate strands behind the anti-essentialist orthodoxy, that understand the significance of variation differently, represent the two major trouble-causing threads in the current debate as I presented it above. For instance, recall that Devitt characterizes essences in individuating terms and then subsequently argues by appeal to explanatory intuitions. Okasha recognizes that essentialist views have included individuating and explanatory aspects, but simply dismisses the latter to focus on the former. The first major lesson I want to salvage from the debate, then, is that outstanding philosophical controversy and questions regarding species individuation should be addressed separately from questions of population thinking-based explanation. Secondly, both topics should be addressed without regard for the essentialism issue. The bulk of my discussion will focus on the individuation issue.

---

17 In fact, Sober criticizes the approach of using variation to discredit essentialism. Aristotle, he noted, was fully aware of the need to accommodate such in his account of biological kinds.

18 This is not to say that variation is not also an explanandum. Rather it becomes an explanandum for another kind of explanans (e.g., microbiological).

Individuating Species Taxa

Devitt and Okasha agree on this much: that we need to answer the question of why organisms belong to the species taxa that they do. As Okasha put it, it cannot be a brute fact that Rover belongs to *Canis familiaris*. What they disagree over is whether to answer the question by appeal to extrinsic or intrinsic essences. It should be clear by now, though, that we can and should ask the question in a different way. I suggest: are intrinsic traits of organisms ever among the individuating features of species taxa? The answer can now be pursued without concern for whether we are meeting any essentialist strictures. I agree with Devitt that a purely extrinsic approach to species individuation isn't feasible, but for non-essentialist reasons. But getting to my own reasons requires exploring some of the issues behind the debate.

If someone asked me whether my pet Rover was a dog, I would direct her attention to Rover's morphological features, i.e., to traits intrinsic to Rover. Of course, there is much more to the story about why Rover is a dog. But surely, one might think, intrinsic traits must be part of it. How could anyone think otherwise? A typical reply is that an organism exactly like Rover in all intrinsic respects, but born of a lineage from a different planet would belong to a different species than *Canis familiaris*. Rover is a dog, a member of *Canis familiaris*, in virtue of being born within the spatiotemporally localized lineage that is *Canis familiaris*.

Surely that can't be the whole story, one wants to say. Both Rover and his hypothesized doppleganger belong to dog-like lineages and that must be a matter of similarity of intrinsic traits. This is where the defender of the lineage conception will appeal to the open-ended nature of what can arise in a lineage. Not everything that

appears in the Canis familiars lineage need look much at all like Rover. Similarly, Rover's doppleganger need not appear in a lineage very much at all like *Canis familiaris*. Furthermore, the appearance of one, or many, novel "varieties" in a lineage does not necessarily create a new lineage. And without a new lineage to belong to, such novel organisms only have the lineages they arose in to belong to. Rover owes his "doghood" to his parents alone, and can claim no contribution from his own intrinsic features.

In reply to this line of reasoning, the champion of intrinsic traits will now demand to know how lineages get distinguished, and rightly so. For the "I owe it all to my parents" approach makes every living thing part of the same lineage going back to Earth's initial organism. One proposal is that lineages are distinguished, at least in part, by the composition of intrinsic traits internal to them. Rover may belong to *Canis familiaris* due to parentage, but *Canis familiaris* is to be distinguished from, say, *Canis lupus* because of differences in traits of the organisms belonging to the two lineages. Hence, it is no accident that Rover has traits found in *Canis familiaris* and not *Canis lupus*.

At this point we come to an odd but intriguing disagreement between proponents of purely extrinsic criteria for species membership and their opponents ("trait proponents"). It concerns the nature and amount of diversity within species. There is a long standing tradition among lineage proponents, going back at least to Hull's classic paper, of denying that there is anything near the amount of requisite unity of traits internal to a species to distinguish them in intrinsic trait-terms. The reason often offered is the apparent failure of biologists to actually come up with adequate lists of

intrinsic traits that genuinely distinguish species (Hull 1965; Sterelny and Griffiths 1999). Indeed, so-called phenetic approaches to species individuation that proceed by seeking to establish similarity metrics are widely perceived as failures, including within many parts of biology (Sterelny and Griffiths 1999; Lewens 2012b).

This idea that there is so much internal diversity is now so taken for granted among some that extreme claims are sometimes made without appeal to biological evidence. Prior to calling exaggeration, Devitt quotes Okasha as claiming that the mechanics of sexual reproduction "ensure an almost unlimited variety in the range of possible genotypes that the members of a sexually reproducing species can exemplify" (Okasha 2002, 196). And this is a common sentiment expressed by many (see Devitt 2008, 370 for more examples).

Devitt goes on to complain that surely this must be an exaggeration, though I think there is more to say than he does about why. It does seem bizarre to claim that there is that much (unlimited!?) diversity internal to species, and it is worth saying something about why. It is usually not said how such diversity is measured. And this is crucially important because depending on how we classify and quantify diversity, an enormous amount of diversity is completely consistent with a large amount of uniformity; or, with low but genuine amounts of uniformity that are nevertheless causally necessary for the maintenance of the diversity as diversity within one species. Taking a wider view of things can help bring these points into focus.

Consider amoebas, pine trees, ants, and elephants. All instances of each share some biological traits—mostly genetic and cellular. If we eliminate amoebas from the group we get even more in common, including some extracellular traits (e.g., multi-

cellularity, sexual reproduction). The list of universal traits continues to increase dramatically as we eliminate types of organisms from our group. Importantly, this is true even when we get to just ants, many species of which have many more polymorphic types internal to them than most other species. Yet, within an ant species, queens will have more biological traits in common with workers than they will with any elephants. Whatever diversity there is within an ant species, it must be compatible with this seemingly obvious fact. Whatever diversity arises within an ant species, it won't be of the kind or degree characterizing our initial group that included amoebas and pines. And whatever diversity there is within elephant species, it certainly doesn't seem like it rises to the level found within species containing the number of polymorphs characteristic of many ant species.

Nor is the above simple armchair pronouncements on biology. As has been recently noted in the philosophical literature, pheneticism is not dead in biology, but alive and well, including for sexual species (Lewens 2012). If this is all so, what would prompt the kind of exaggerated assertions about intraspecific diversity so common in philosophy of biology? One reason may be fear of essentialism. Species anti-essentialism has always accompanied philosophical appreciation for the importance of variation in biology. But "the importance of variation" can mean different things, as we see in the Mayr-Sober strand of anti-essentialism. Its "importance" need not be compromised by admitting that it has its limits within species.

But if we can agree that the essentialism issue need no longer be a concern, we can approach the issue of intra-specific diversity afresh. For clearly there are puzzles here. It is true that various phenetic projects of the past have not attained their aims.

There aren't always simple ways of saying how intrinsic traits distinguish closely related species. And new species do arise, gradually, from old ones. Wilson, Barker, and Brigandt (2007) nicely bring out the "intrinsic heterogeneity" of species and the need to capture their simultaneous "integrity" and "flexibility". I think this speaks to the fact that we need more tools for talking about diversity. Does diversity come in kinds or merely degrees? What kind of diversity matters to species individuation and what doesn't? In the next section I would like to offer a way of talking about variation that may help the debate.

Types of Variation

Social insects such as ants are well-known, paradigmatic cases of species with several polymorphs. But the kind of diversity in an ant colony due to characteristics fundamentally distinguishing queens and soldiers seems different in kind from the sort of diversity that distinguishes individual cheetahs by each's top sprinting speed. Roughly, one is diversity in type whereas the other is diversity on a continuous scale. To distinguish these kinds of diversity, I introduce the categories of stable variation and schmear variation. The category of stable variation is meant to cover reliably recurring variations that may often be thought of as sub-types of some larger classificatory unit. I have in mind at least polymorphisms in species, developmental stages of organisms, and alternating generations. Other kinds of phenomena may fall under this category as well. The category of schmear variation is meant to cover at least the ongoing, often random, generation of gradual variation that can be well illustrated by traits measurable on continuous scales (e.g., height). But it need not be limited to such. I take both kinds of variation to intersect, but not necessarily coincide, with heritable variation.

We can use these categories to think more clearly about what threat variation poses to individuation of species via intrinsic traits. For instance, suppose we said that to be an ant of a certain species is to be a schmear variant of one of its stable variants, where these are a matter of intrinsic traits? Or consider stereotypical traits of creatures that we might want to use to individuate their species. Suppose someone proposed to individuate *Giraffa camelopardalis* by its members' long necks. There will be a significant amount of schmear variation with respect to the length of their necks. Does this kind of variation rule neck length out as an individuating feature? Not necessarily. We could propose, instead, that to be a member of *Giraffa camelopardalis* is to be a schmear variant with a neck length falling within a specified range. The fact that all giraffe necks fall within a certain range is most likely a function of the history of selection pressures they faced. And creatures with necks falling outside of that range also have compensatory differences elsewhere in their anatomies and physiologies. So it may be quite informative to characterize them in that way.

Another proposal from this way of thinking about variation is that, at least in some cases, once schmear variation reaches a certain point, we may have a novel stable variant. Schmear variation can only go so far before it turns into something different and requires accompanying changes elsewhere in the system to accommodate its structural demands or facilitate its new functions. If giraffes began to face selection pressures for even longer necks than their current upper limits, they would need to evolve further skeletal and physiological changes to accommodate them, eventually resulting in a new species (of giraffe?) or perhaps a new polymorph within the same species.

I have used fairly simple examples here to illustrate advantages of my approach to variation for making it a more tractable issue to think about. But I see no reason in principle why my categories couldn't be extended to talk about more complex or non-obvious examples of variation throughout the biological realm. Perhaps further categories of variation are needed to cover the different kinds. But in offering a way to think about them so that variation as we find it need not be the death knell of intrinsic individuation of species taxa, I present a challenge to pure lineage conceptions of species. It is: be more specific and find a kind of variation that cannot be accommodated by what I have offered here, and in a way that genuinely threatens intrinsic individuation. Otherwise, appeals to variation should no longer count against it. Stable and schmear variation, of the sort that actually obtain in Earthly biology, can both be accommodated by intrinsic trait approaches to species individuation in ways motivated by evolutionary and developmental biology.

That is my claim, anyway. I realize that a stronger case would need to be made, one that includes a proposal for exactly how intrinsic traits should figure in the individuation of (perhaps only some) species. What motivates re-examination of the issue are strong, yet non-essentialist reasons why purely lineage-based views of species are wrong. The main reason is that they are under informative. They do not distinguish species from other biological lineages (e.g., genes, cells, etc.). Nor do they distinguish the species we are familiar with from possible species: those that are even more uniform than most Earthly ones, or ones that are far more internally diverse. Purely lineage-based conceptions of species fail to distinguish them from lineages with some otherwise very different properties and dynamics—both actual and possible. They do not

acknowledge how rich "lineage space" can be in principle, or the resultant need to locate our biological world in it.[19]

My proposal for how to start thinking more about the interplay between uniformity and diversity within and across species may be on the wrong track. But I hope to have shown that the implications of variation for the individuation of species is a more complicated issue than often recognized, and that it needs to be addressed without distractions from the specter of essentialism.

Explanatory Issues

There is more to say about the Mayr-Sober/explanatory thread in the way of lessons than questions. First is to re-emphasize the genuinely explanatory focus of this strand. Much of the neo-essentialist revival has tended to focus on the individuation question. Devitt simply claims agreement with Sober's substantive points and that his essentialism need not have any quarrel with him (371-372). This move underestimates the significance of the Mayr-Sober line of thought that seeks to gain reflective understanding about the sea change in our scientific understanding of the biological realm due to Darwin and Darwin-inspired developments.

There is no doubt that the introduction and maturation of "population thinking" as noticed by Mayr, and better developed by Sober, constitutes a profound change and improvement in our understanding of biology that is worth reflecting on. It has many implications, including ones that genuinely challenge what our everyday forms of

---

19 This is true of de Queiroz's (1999) otherwise wide-ranging defense of the lineage conception of species. It is also the fatal flaw in Matthen's (2013; 2009; Ereshefsy and Matthen 2005) repeated attempts to fault HPC-kind approaches to species for failing to adequately deal with what I call stable variation. Matthen cannot account for why stable variations within species are not nearly as diverse as they could conceivably be, nor why key differences in otherwise very similar cross-species polymorphs render them "incompatible" as within-species polymorphs.

generalizing judgment about biology lead us to believe. In fact, this may be one better way of stating the change. Aristotelian biology enshrined, and gave systematic backing to, our everyday forms of judgment that lead to what Sober called the "natural state model" of species, and which Michael Thompson (2008, 2004, 1995) has recently tried to defend. In contrast, just as physics and chemistry reached certain milestones in their development that forced us to give up in our theoretical beliefs what we find incredibly difficult to give up in our common sense judgments (the sun rises and sets), population thinking forces us to give up in our biological theories ideas that we may still find natural in everyday life.

Another way to appreciate population thinking is to see how it instituted a new form of explanation unknown to physics and chemistry. There are two ways to understand Kant's (1790) claim that biology would never have its Newton. One is that our unsystematized teleological understanding of biological entities would never be reduced or systematically explained. Another, more specific way, is that it would never happen in a Newton-like way, i.e., with a few basic, universal laws governing the dynamics of all biological entities—perhaps ones that could then easily be reduced to the laws of physics. The advent of population thinking refutes Kant on the first interpretation, but does so while retaining principles specific to biology and unlike any that had been found in physics or chemistry.[20]

From this perspective, Mayr and Sober were too broad brushed in some ways, and not broad enough in others. Claiming outright that population thinking had

---

20 Caveat: Principles of populations first discovered in biology and not present in most basic physical systems. I do not mean to imply that there are no non-biological domains to which population thinking does not extend.

replaced typological and essentialist thinking was overly general and too ambitious. At the same time, focusing on figures of the past (Plato and Aristotle) also encouraged neglect over other ways in which population thinking contributes to insights distinctive to biology. Furthermore, population thinking does not cover species only. As Peter Godfrey-Smith's Darwinian Populations and Natural Selection has firmly established, principles of population thinking extend throughout the many levels and regions of biology. They apply to different types of things (genes, cells, etc.). The reach of population thinking is profound. But it doesn't rule out grouping things together on the basis of shared properties.

Finally, while my position must reject Walsh's neo-essentialism, I do think his critique of Sober indicates something important. One is to be cautious about exuberant claims about breaks with the past and over reliance on one innovation in biological thinking. There is much that remains unknown about biology and further developments may bear some resemblance to ideas from the past. The bottom line here is that understanding the significance of population thinking in biology remains an important project that requires more nuance than the anti-essentialist banner allows.

The Essentialist Impulse and "Psychological Essentialism"

Any position accusing a debate of fundamental confusion over an ancient concept of Western philosophy owes some sort of diagnosis or explanation for the situation. How could such smart people be so deluded for so long? One key to the diagnosis stems from the observation that essentialist positions seem to survive profound changes in what is considered admissible approaches to individuating entities and explaining

generalizations containing them. The profoundest such change, of course, was inaugurated by the scientific revolution and its rejection of Aristotelian substantial forms, teleology, etc.

My suggestion is that there has been a mistake about what is common across the changes. What survives such profound changes is not a stable metaphysical position or thesis (there are essences and *this* is what they are), but something closer to a regulative ideal that gets mistaken for such. That regulative ideal enjoins us to connect our individuating and explanatory practices as closely as possible. I'm more inclined to call it a regulative *impulse* or *drive* however, because I don't think it has much more determinate content than that. This is why there can be such disagreement over what counts as a "truly" essentialist position with respect to F's. What Aristotle correctly recognized is that there *is* interdependency in our individuating and explanatory practices.[21] Successful explanation depends on picking out the right explananda and explanantia, and in the right sort of way. And the entities worth positing and being precise about are generally those that we identify as having explanatory power. What is right about anti-essentialist resistance is both the willingness to recognize when nature resists our desire to make individuation and explanation as simultaneously simple, unified, and powerful as possible *and* the refusal to turn our regulative drives into contentful metaphysical theses.

Part of the explanation for the confusion is surely Aristotle's genius for recognizing an intellectual drive and articulating it into an explicit and worked out thesis. That set a powerful precedent for recognizing essentialism as a metaphysical or

---

21 Charles 2003 brings this out well.

theoretical doctrine. I'll bet there is also some psychological tendency to reify, that crystallizes around the use of "essence". What is potentially most interesting to my mind, however, is the connection between my speculations here and the psychological research program on native "essentialism".

Psychologists have identified a tendency toward a suite of judgments in very young children that gets them labeled "essentialist" (Gelman 2003). Anthropologists have documented that at least some of these tendencies are culturally universal (see Atran 1999 and Atran et al, 1997). If my line of thought here is correct, it is wrong to identify this psychological tendency with a universal and coherent philosophical position (for there isn't one). Fortunately, there is precedent in the literature for resisting this kind of identification. Michael Strevens (2000) argued that all the extant evidence was consistent and better accounted for without the assumption of a psychological posit of essences. While at least some psychologists have claimed to successfully respond to Strevens, my approach breathes new life into a Strevens-like project of re-examining and potentially re-interpreting the data (Gelman 2003). It could also suggest alternative hypotheses to test in this research area as well. Finally, it raises questions about how the *mature* intellectual drive I posit is related to the innate tendency toward "essentialistic" judgments observed in children. It also asks us to consider whether the essentialist impulse can be discerned in other intellectual traditions. Further development of these admittedly rough ideas will have to await another occasion.

Conclusion

The species essentialism issue is irreparably broken. It not only now fails to have the unity required for a solvable debate, the original anti-essentialist consensus lacked the unity that could be appealed to for restoration. Intraspecific variation is crucial to understanding biology, but its importance has been misapplied and even overstated in some ways. It has different implications for how to individuate species taxa and how to understand the rise of new forms of explanation in biology. These two issues should be distinguished from each other and treated without concern for essences. There remains important work to be done on each. Further, coming to understand why the essentialism issue really is irredeemably confused also holds promise for shedding new light on the idea that humans are native "essentialists".

MORE BARK THAN BITE: THE INNOCUOUS IMPLICATIONS OF THE EVOLUTIONARY DENIAL OF HUMAN NATURE

Introduction

If you accept the evolutionary account of how humans came to be, and you understand it correctly, then you will reject human nature as a myth of a bygone era. That is the influential claim of many experts in evolutionary theory, including both biologists and philosophers of biology. The influential biologist-philosopher Michael Ghiselin declares: "What does evolution teach us about human nature? It tells us that human nature is a superstition" (1997, 1).[22] It is a claim that seems radical and revolutionary on its face. It sounds like it should have many important consequences. But it doesn't. At least, any sense in which it is true does not.

The first part of this paper examines the arguments meant to establish the absence of any human nature. "Human nature" is a vague and amorphous term.[23] Consequently, recent authors have realized, and argued, that there may be perfectly legitimate senses attached to it, which the evolutionary arguments don't touch (see especially Machery 2008; Samuels 2012). I applaud the aims of these efforts, but think they proceed under the guidance of a too quick acceptance of the supposedly sound evolutionary arguments. While there are some sound parts of these arguments, suitably reconstructed, there is a lot of chaff surrounding them that needs to be separated and

---

22 The mythical status of human nature is equally forcefully asserted by David Buller in his influential attack on evolutionary psychology (2006, 480). David Hull's 1986 "On Human Nature" is the locus classicus for this kind of view, which has been re-affirmed most recently by Tim Lewens (2012). See the introductory texts by Sterelny and Griffiths (1999) and Godfrey-Smith (2014) for a sense of the widespread impact on the field.

23 But good illustrations of well-known projects that would be threatened by the arguments considered here are on display in Pinker (2003) and Tomasello (2001).

dispensed with. Doing so can contribute to illuminating how the project of articulating a scientifically sound conception of human nature should proceed.

It will also help us figure out what can be inferred from the rejection of human nature. This is the task of the second part of this paper. To take one example, the nativist-empiricist controversy is often associated with the investigation of what human nature is like (see, e.g., Prinz 2012). One might easily think that the evolutionary rejection of human nature should favor the empiricist-side of this debate, which also sometimes characterizes itself as showing that there is no such thing as human nature (ibid). As it turns out, though, the evolutionary rejection of human nature is fully compatible with the strongest of nativist theories of human psychology. There are other major consequences that the advocates of the evolutionary arguments have thought followed—for normative ethical and political theory; for the character of the human sciences; and for kinship with other philosophical movements. As I'll show, none of these supposed implications follow. To put it most provocatively, even if there is no human nature, much science and philosophy concerned with 'human nature-y' topics may proceed as though there is, with minimal risk of any serious error.

Evaluating the Arguments

As noted, "human nature" has many senses and uses. Rather than try to pin them down, I find it more effective to proceed by examining the arguments that have been offered against it. The main arguments come in two flavors, one more theoretical and one more empirical. The first type relies on a general anti-essentialism about biological

species thought to be at the heart of evolutionary theory. The second appeals to actual diversity found in the species to which humans belong. Here are schematic summaries.

Theoretical Argument

Incompatibility Evolutionary theory is incompatible with the existence of species essences.

Therefore,

~Essence *Homo sapiens* has no essence shared by each of its member organisms (Intermediate Conclusion)

Coverage All humans belong to *Homo sapiens*.

Bridge Human Nature = common essence due to membership in *H. sapiens*.

Therefore,

~Human Nature There is no human nature.

David Hull gives succinct expression to an instance of this type of argument:

> "If species are interpreted as historical entities, then particular organisms belong in a particular species because they are part of that genealogical nexus, not because they possess any essential traits. No species has an essence in this sense. Hence there is no such thing as human nature" (1978, 358).

In what follows, I propose to grant **Incompatibility**, or at least a suitably interpreted and explained version of it which I provide in the next section. My aim is to show how little this weighty premise gets deniers of human nature. I do this in two ways. First, by cutting **Incompatibility** down to size through constructing a conception of human nature that is a candidate for confirmation by the human sciences (i.e., has not yet been ruled out by current evidence) and showing why it does not violate anti-essentialist

constraints. But I also insist on the necessity of, and substantive work being done in the argument by **Coverage** and **Bridge**. These latter two premises often slip by as implicit; or as so obvious as to not need argument; or as the only acceptable assumptions of a suitably naturalistic philosophy. But matters are not so simple. It turns out that non-essentialist conceptions of human nature can be more substantial than deniers or deflaters of human nature suppose.

Empirical Argument

Uniformity	If there were a human nature, it would consist of traits held by humans for which there could be little to no variation across humans
Diversity	There are no human traits for which there is little to no variation across humans.
	Therefore,

~Human Nature

We will look at more detailed arguments of this kind later, but David Buller expresses the thinking behind it thus:

> "…any psychological universals we might happen to discover—if we were to discover any at all—are temporally contingent. Today's universals were not yesterday's universals, and today's universals may be possessed by only a fraction of our species, or even extinguished altogether, tomorrow. Thus, any psychological universals we might happen to discover in the present would not characterize Homo sapiens per se, but would only characterize our species at this particular moment in evolutionary time" (2006, 477).

In examining this form of argument, I again grant the first premise (**Uniformity**). My focus will be on **Diversity**, and I show the arguments for it to be lacking. I also consider modifications of **Diversity** that also turn out to be indefensible. Or, at least not

defensible via the resources of evolutionary theory, which is the primary impetus behind the denials of human nature considered here.

Along the way, I flag and reject what I think of as "filler" arguments that have been part of the mix but which have even less force than those above, serving only to distract. By the end of this part of the paper, then, we should still be ready to affirm the incompatibility of evolutionary theory and conceptions of human nature containing essentialist commitments. But we should also be more antecedently skeptical of the power of that fact to imply broader conclusions. In the second part of the paper, I turn to affirming that skepticism for specific consequences that have been thought to follow. For now, I turn to the explication and examination of each of the above arguments in turn.

Arguing via Evolutionary Theory

The linchpin of the theoretical argument above is **Incompatibility**—the idea that evolutionary theory and biological species essences just don't mix. How is this incompatibility supposed to arise? Here, too, the question is vexed, also because of vagueness and varying uses of the key concepts. In another paper, I argue for eliminating the essentialism issue from debates about the metaphysics of species.[24] While I won't depend on that conclusion here, I will lean on a lesson I draw from that debate. It is that we should identify more specific commitments usually associated with essentialism that mature evolutionary theory (i.e., not Darwin's ideas alone) cannot

24 See my "What to Salvage from the Species Essentialism Debate"

allow, rather than try to argue that it forbids essences per se.[25] This, I think, is all the "anti-essentialism" we can garner from evolutionary theory, as well as all that we need. As a consequence, theories of human nature need only avoid those specific essentialist commitments in order to be compatible with evolutionary theory. In order to get to these commitments, though, we need to examine some common elements of anti-essentialist arguments that do not work.

There are two types of evolutionary anti-essentialist arguments to consider. One, alluded to in the quote by Hull above, attempts to reclassify the species category so that none of its instances are the kind of thing that can have essences. The other makes various appeals to the role of intra-specific variation in evolution. As I show, only one of the ways of appealing to variation's role leads to an appreciation for its incompatibility with certain essentialist commitments.

Ontological Category — Some argue that evolutionary theory recommends reclassifying the species category (Hull 1978; Ghiselin 1997). Before Darwin, species were understood to be kinds. Now they are understood as individuals. And individuals, it goes, do not possess essences, at least not in the same way as kinds do. In brief, all of the whale sharks do not share a common essence because the species to which they belong is not the sort of thing to possess an essence or to bestow one on them.

But even if we grant individualism about the species category and the metaphysics of individuals the argument assumes, the essentialist impulse can adapt.

---

25 This is more or less the route taken by the most careful participants in the species essentialism debate, anyway. The part of my paper that I don't rely on here argues that even this way of proceeding cannot save the essentialism arguments, on either side, and so cannot save the debate as a whole. If the reader comes to suspect that the essentialism considered here is weaker than the essentialism that evolutionary theory can justifiably reject, I recommend my other paper. The upshot of that paper is that such supposed essentialist positions are will-o-the-wisps that need not concern what genuine issues remain in the metaphysics of species.

On this view, whale sharks now relate to their species as parts relate to a whole. Whale sharks may all still share a common essence, though. It is generally fallacious to deduce the lack of a property in a whole to a lack of a property in each of its parts. And there appears to be no reasons to think that essences are a special exception.[26]

I raise the ontological category issue because two of the most influential and vociferous opponents of human nature, David Hull and Michael Ghiselin, are also two of the most influential advocates of species individualism, and Hull first raises the human nature issue at the end of his most widely cited defense of the individualist thesis (which I quoted from above).[27] But there are other arguments as well, which are more important, and I think it is crucial not to get this issue mixed up in the human nature debate.

So the maneuver of switching ontological categories will not do the trick. The essentialism issue can just be recast. But there is a family of considerations of a different sort that may underly both the ontological shift and the anti-essentialism. Their common feature is that they all appeal, somehow, to how variation within species is understood, though this contributes to an illusion of more unity to this range of considerations than there really is. When pulled apart it becomes easier to see that they span a wide range of capacity to actually undermine essentialism.

Variation — The historical element to the theory of evolution—the claim that all organisms are related by descent from one, or a few, initial organisms—demonstrates that biological relatedness is compatible with diachronic variation in biological traits of

---

26 Both Devitt (2008) and Boyd (1999) argue what is basically the same point.

27 Buller (2005) also uses this argument has his main one against human nature.

an extraordinary amount. But unless we are willing to assert that the entire tree of life is the only species, the sheer fact of this kind of variation should not undercut species essentialism. For as long as we distinguish different species on the tree of life, we can ask whether they have essences. Variation must play a different kind of role.

Variation has an indispensable function in the explanatory element of the theory of evolution—the idea that the historical facts mentioned above can be explained via the process of natural selection. That process requires a specific kind of variation. Namely, variation in heritable traits that contributes to patterns of differential reproduction in the entities (e.g., organisms) that possess them. As David Hull says, without this kind of variation, evolution (or at least evolution via natural selection) would "grind to a halt" (1986, 3). While true, this point is not itself reason to reject essentialism. There is no law prescribing that evolution must continue, and no reasonable essentialist position need concern itself with whether or not it does. Still, proper understanding of the role of variation in natural selection does require reorientation away from some tempting ways of conceiving of the biological realm, including ones often characterized as "essentialist".

Historically, essentialist positions generally have both definitional and explanatory commitments.[28] So a conception of human nature avoiding both kinds of commitments may be certified as essentialism-free. In the following sections, I will examine the issues of how the kind of variation required for natural selection is

28 There is not an explicit consensus on this in the essentialism literature, either generally or with respect to the species issue. But, clearly, a conception of human nature that avoids both kinds of essentialist commitments need not concern itself with the outcome of this issue. At best, the outcome of the issue would allow a non-essentialist conception of human nature more leeway than I am allowing here.

supposed to reorient species individuation and our explanatory framework for traits commonly found in a species.

Variation and Species Individuation—When Hilary Putnam helped revitalize "essentialism" in "Is Semantics Possible?" and "The Meaning of "Meaning"", he presented a certain picture of how we pick out kinds with "hidden" essences. We can encounter something that we treat as a sample of a kind and then say that it bears the same substance relation to other instances of the kind. We do this even though we do not know, or may even initially be mistaken about, what properties determine the same substance relation for the sampled kind. This is the story he told about water. The properties determining the same substance relation between all possible samples of water may be thought of as its essence.

As with water, one is tempted, so with species. One encounters a wolf for the first time and intends to count it as of a kind with all other organisms that bear the same substance or "same species" relation to it. This is a tempting picture indeed, and may be an apt description (or idealized description) for how naturalists have classified species pre- and even post- Darwin. But the historical biological facts and the facts about natural selection mentioned above suggest a different approach to grouping organisms.

If natural selection is a or even the driving force in the history of life, we ought to classify biological entities in a way that helps us track its movement through that history.[29] As described above, natural selection requires organisms that vary with respect to heritable traits. To keep track of natural selection, then, we need to keep track of such variation. This kind of variation is a property of collections of organisms. And

29 My way of framing the issues here has been significantly influenced by Godfrey-Smith's (2009).

changes in that variation are tracked by changes in those collections. These collections are called populations, and populations are the primary locus of evolutionary change. As the variation that matters is variation in heritable traits, we need to track relations of inheritance—we need to track relations of reproduction. Populations, to the rough approximation necessary for our purposes, are reproductive communities, or communities competing over the same resources (including those necessary for reproduction). Reproductive communities give rise to lineages of descent.

To track the influence of natural selection, the primary groupings we need to track are Darwinian populations and lineages. From the point of view of natural selection, then, employing Putnam's method upon encountering a wolf is the wrong way to start. What we should be interested in is not: what same substance relation would group this wolf with other organisms? At least not where the same substance relation is modeled on that of the microstructural approach to chemical kinds like water. Insofar as that approach recommends picking out a simple, uniform genetic profile from a wolf to determine wolf-hood, it would fail to account for wolf sexual dimorphism.

Instead, then, we should ask: what Darwinian population and lineage does this creature belong to? And given that Darwinian populations are characterized by some degree of diversity, we cannot presume ahead of time what sorts of similarities we should expect to find across all (and only) members of the relevant Darwinian population. Nor should we run the risk of limiting, ahead of time, what sorts of novelties may arise in populations or lineages by seeking to delimit them from the point of view of a Putnam-style search for the "same substance" on the basis of extant samples.

The upshot of these considerations is that, from the point of view of Darwinian evolution, organisms should be grouped together on the basis of population and lineage membership, not a grouping criterion picking out a simple set of duplicable microstructures determining something like a Putnam-esque "same substance" relation for species members. This is not to say that all and only wolves won't have some "microstructural" things in common. But, for instance, the Darwinian population that they belong to will depend on some microstructural (and other) differences between many of its member wolves (e.g., ones necessary for sexual dimorphism).

Consequences for Human Nature

The consequences of the above for human nature may now be stated. If the biological taxon *H. sapiens* is individuated according to lineage-population criteria, then our concept of human nature should not propose a contrary way of individuating the *H. sapiens* taxon. It should not propose membership criteria that include organisms that the lineage criteria exclude nor include organisms that the lineage criteria exclude.

One easy way for a conception of human nature to avoid conflicting individuating criteria is by not entering the individuation game at all. It can simply refuse to be about species classification. This may seem overly revisionary. Surely, one might think, the notion of human nature was always meant to help us figure out what counts as human. But I think the charge of revisionism cuts more strongly in the opposite direction. It is a revisionist interpretation of the aims of traditional interest in human nature to suppose that it was ever meant to compete with the aims of the lineage approach to delineating the H. sapiens taxon. It may turn out that the class of organisms of interest to human nature theorists is a sub-population or sub-lineage of *H. sapiens*.

As I'll discuss more once we have the explanatory element of essentialism on the table, the foregoing shows why **Coverage** and **Bridge** are so much more important to the theoretical argument against human nature than usually acknowledged. There appears to be an implicit and challengeable assumption of a theoretical identity claim between the class of things usually targeted by human nature theorists and the members of the lineage that biologists designate as H. sapiens.

For now, let me note an important limitation of this way of accepting the lessons of species anti-essentialism. The argument just given does not rely on, nor imply, a principle according to which all heritable traits in a species either do or even can vary across generations. In fact, it is fully consistent with a principle according to which the variation required for natural selection is dependent on the existence of other invariant traits.

Variation and Explanation

The absence of essences qua same-substance determiners is a problem for essentialist modes of explanation in biology as well. Suppose we have two organisms that are supposed to be conspecific due to a presumed same-substance relation. Then, if they significantly differ, it becomes a special problem to explain why. One organism is a "normal" instance of the kind and the other is, somehow, not. Hence, we get asymmetries in form of explanation for their traits:

Essentialist Tiger Claw Explanation

Kind Membership Organism X has claws because it is a tiger.

Disturbance	Organism Y does not have claws because, although it is a tiger, there was an interference preventing it from possessing claws.

The essentialist mode of explanation for why two organisms in the same species differ with respect to certain traits employs an asymmetry in form of explanation. One explanation appeals solely to what is “natural" to the organism in virtue of the kind it belongs to. The other appeals to factors purely extrinsic to the organism to explain its trait.

In contrast, we now know that there are efficient causal mechanisms, both internal and external to an organism, which, in combination, are wholly sufficient to explain the traits that any two organisms have. Differences between organisms are to be explained by differences in the developmental history of each organism. Gone is any explanation of difference by form of explanation. It is replaced by explanation in terms of locating differences in the chains of efficient causation that constitute development. Now we have:

Modern Biological Tiger Claw Explanation

Development	Organism X has claws because of the developmental trajectory of its genetic material embedded in its particular environment.
Development	Organism Y does not have claws because of the developmental trajectory of its genetic material embedded in its particular environment.

Organisms may differ over whether the main factors with respect to a given trait were primarily genetic or more environmental or close to an equal share. The point is that there is not a difference in fundamental type of explanation.

Of course, we also need an explanation for why tigers tend to have claws. But there is no appeal in modern biology to anything like a tiger-substance that has claw possession as part of its "nature". Rather, modern biology appeals to selection histories within populations to explain how certain kinds of claws came to be prevalent, or even fixated, in tiger lineages. We are also learning more and more about how we will have to incorporate facts about regularities of robustness in developmental pathways to fully adequately explain the historical patterns of selection generating tiger-claws.[30] But this is best understood as ongoing discovery of facts about development, and not any kind of return to abandoned forms of explanation.

The foregoing is meant to arrive, by a slightly different route, at much the same point that Eliot Sober argues for in his 1980 paper, "Evolution, Population Thinking, and Essentialism," which is a refinement and development of Ernst Mayr's (159/1975) distinction between "population thinking" and "typological thinking". There he makes the case that species essentialism is committed to a "natural state" model according to which a type is identified as the norm in a species and all intraspecific diversity is understood as deviation from the type. In contrast, population thinking understands any norm-designated type as just one variant among others. It may be most prevalent. And its prevalence may have significant consequence for population dynamics. But that

30 See Wagner (2014) for a good recent summary on this issue.

significance has to do with general principles that have been discovered in population biology, not with anything like a telic-function associated with the prevalent type.

Human Nature without Essences

For a conception of human nature to be anti-essentialism compatible, then, it must not propose individuation criteria for Homo sapiens that rule out or ignore its character as a lineage. It also cannot play the role of a normative standard in relation to which we understand either (human) psychological generalizations or individual (human) psychologies.

But the above constraints leave plenty of leeway for a conception of human nature. It may still have both descriptive and explanatory features. In particular, "human nature" may pick out the set of psychological features (if any) that evolved to fixation around the onset of what has been called "behavioral modernity" roughly 40,000 years ago (Nowell 2010). Or perhaps the subset of those features that have remained at fixation since that time. Call this latter conception the inertial behavioral modernity conception of human nature (IBM).

The IBM conception does what a theoretical proposal should. That there is something to it is a good empirical bet. The proposed set is extremely likely to be far from empty. It is likely to contain features that help explain the evolution and character of early human cultures, and perhaps even broad features of contemporary cultures. Yet it does not impose on science's ability to surprise us with discoveries about it. Perhaps science will tell us that if evolved universal or species-typical psychological traits are important to our conception of human nature, then we should include features that appeared later than whatever the onset of behavioral modernity turns out to be. IBM

doesn't, of course, rule out the evolution of any number of further psychological traits in the *Homo sapiens* lineage.

Any resistance to IBM from our theoretical argument cannot come by way of **Incompatibility**. Rather, it can only come from Bridge, the principle equating human nature with a common essence shared by members of *Homo sapiens*. There are two parts to **Bridge** to consider. The idea that human nature must be an essence and the idea that, whatever human nature is, it must be common to all members of *Homo sapiens*.

The idea that human nature, if it is to exist, must be an essence of some sort or another has multiple problems. In this context, it is important that this is an extra-evolutionary assumption. That might be justified if it was obvious that all those, especially in the human sciences, who work on human nature thought of it as an essence of some sort. But this is not a good interpretation of the average scientist of human nature, or even past philosophers (e.g., Hume[31]). There is no indication that most psychologists think of themselves as investigating the human essence in anything like a technical philosophical sense, or of espousing the essentialist tenets that are genuinely incompatible with evolutionary theory.

Next is the assumption that the class of creatures with human nature must be co-extensive with the class of creatures in Homo sapiens. Perhaps this is encouraged by the **Coverage** premise, but it should be recognized that it is not logically required by it. **Coverage** is compatible with the set of humans being a proper subset of the set of creatures in *Homo sapiens*. Given the lineage approach to individuating species, there

31 See Samuels 2012 for a good argument in support of this claim.

need not be any assumption that "humans" was ever meant to pick out precisely what the lineage approach to species will.

To put this point a bit differently, the lineage approach to species does not rule out the idea of different kinds or, maybe better, natural "joints" arising over the course of the lineage. In fact, to recognize the emergence of behavioral modernity just is to recognize the arrival of such a new joint. Insofar as the various fields in the human sciences are actually tracking something when they claim to be studying "human nature", it may be the "joint" that arose with behavioral modernity and continues today. Or a different such joint that future scientific work may discover. The point is that nothing in the theoretical argument rules this out. Insofar as invocations of human nature do partially track something, there is no need to antecedently rule that out by saying that it must track something among all the creatures picked out on a lineage-based individuation of Homo sapiens. Whether it does pick out anything should be more of an empirical claim.

The IBM conception of human nature survives the theoretical argument. It is subject to theoretical and empirical refinement or even replacement. An alternative may be desirable for any number of reasons. I neither fully endorse nor reject it here. But the idea that it or any reasonable alternative is ruled out in principle by evolutionary theory is simply false. However, it does rely on a conjecture that there is bound to be quite a bit of psychological uniformity over time in human populations. So it is time to examine the more empirically oriented argument against this conjecture.

The Empirical Argument

The dialectic above deals with the claim that human nature is somehow incompatible with evolutionary theory in principle. I argued that this is true only for certain highly theoretically loaded conceptions of human nature. Ones that it is uncharitable to attribute to most working scientists who think of themselves as studying human nature. In this section, I consider an argument aimed at any reasonable conception of human nature worth the name. It asserts that, as an empirical matter, there is more human diversity than any theory of human nature can abide. I argue that the empirical facts do not show this and that the idea that they do is motivated by extra-empirical assumptions.

A version of this argument has been affirmed most recently by Tim Lewens in "Human Nature—the Very Idea" (2012). But it is a theme in the writings of David Hull and David Buller as well. They argue that the idea that there is any significant amount of uniformity, or even species-typicality, in human beings is an illusion due to the present moment of human beings' evolution. Buller calls the tendency a "temporal provincialism" (2005, 477-8). In the following arguments, then, all claims about human diversity should be understood to be temporally unrestricted. I show that the argument that there is not sufficient temporally unrestricted uniformity in human characteristics for a robust human nature does not succeed.

Recall the "empirical argument" outlined at the beginning:

Empirical Argument

Uniformity If there were a human nature, it would consist of traits held by humans for which there could be little to no variation across humans

Diversity There are no human traits for which there is little to no variation across humans.

Therefore,

~Human Nature

In what follows, I will leave aside **Uniformity**. While contestable, I think it better discussed in the course of constructing a positive conception of human nature. Diversity is the main claim of the human nature deniers and I think it can easily, and illuminatingly, be met.

In addressing **Diversity**, note first that it involves a negative existential. Since human traits number at least in the thousands, and there are limited resources, **Diversity** cannot be confirmed by direct observation. It must be backed up by inductive support. There are certainly some traits that are not universal. Not every adult human being has 20/20 vision. But we still need a reason for why we have a sufficient basis from which to project **Diversity**. Given that various scientific fields continue to discover new types of human traits (often quite surprising), it seems hasty to conclude that the basis for projecting from the diversity we do know of is that it is just so observably ubiquitous. We need reason for why we should expect that novel kinds of traits that take extensive empirical investigation to discover should not be uniform across humans as well. I think that the role of variation in natural selection is doing double duty here. It leads some to think that there must be variation everywhere. But, as I argued above, that is not required by the theory of natural selection. That some variation is needed does not entail that there is variation in all traits.

Appreciating the theoretical motivations for why **Diversity** might be thought to be an empirically safe bet is important. Because **Diversity** is just plain false—and seemingly obviously so. All humans have: bones, sensory neurons, motor neurons, blood vessels, etc. Accommodating these basic facts about humans demands revising Diversity. Below are two modified descendants of **Diversity**, versions of which have played a role, sometimes just implicit, in arguments in the anti-human nature literature.

Distinctiveness There are no distinctively human traits for which there is little to no variation across humans.

Propriety There are no human traits for which there is little to no variation across humans and which are proper candidates for human nature.

I will address both **Distinctiveness** and **Propriety** in turn.

Distinctiveness demands that the existence of human nature entail that of all traits unique to humans, at least one has insignificant amounts of variation across humans. Why should this demand be recognized, though? Suppose it turned out that human nature was continuous with some other nature, perhaps chimpanzee nature or primate nature or even mammalian nature. If this were indeed so, we would be justified in saying something like, "human nature is nothing over and above primate nature". But that would be very different than saying something like, "well, it turns out that humans have no nature". There is a huge chasm between the state of affairs in which humans have no nature at all and the one in which they simply share natures with other biological beings. Claiming there is no distinctive human nature is importantly different

from claiming that there is no human nature at all, and the dialectic in the literature should better reflect that.

There is more to be said to directly confront **Distinctiveness**. What is the empirical evidence for it supposed to be? There are many kinds of traits that are unique to humans, under many different categories: genetic material, developmental and ageing processes, physiology, morphology, psychology, and behavior. Are we really to believe that amidst all this uniqueness there is no significant uniformity? I do not find convincing arguments to this effect, but only bald assertions that do not demonstrate any sort of biological consensus on the issue.[32] Let me discuss, though, an example of a type of argument that may be taken to suggest otherwise. It appeals to actual examples of human diversity and tries to infer broader conclusions about its extent.

Polytypicality Projection

Many arguments of this flavor appeal to the polytypic nature of certain traits within human populations. Such arguments are potentially powerful because they do not appeal to slight variations or gradations that might suggest that the relevant variation is largely ephemeral or insignificant. Rather, they show that there are stable, causally salient variations in evolutionarily significant traits. In a specific case of this argument type, David Hull (1986, 5-6) appeals to human blood type which is certainly polytypic. Apparently more than 30 blood groups are recognized by scientists. Most people will be familiar, though, with the A, B, AB, and O groups (and each's +/-

32 Buller (2006) is a good example of repeated assertions to this effect and in a tone that treats it as either obvious or part of some well established scientific consensus. But there is no citation of literature establishing such a consensus. See Machery and Barrett (2006) for an important critique of Buller's misinterpretation of the literature he criticizes and which also presses him on his exaggerated claims concerning diversity. Devitt (2008) also presses this point against anti-essentialist arguments. As my earlier arguments should indicate, I disagree with him over whether the exaggerations of variation are relevant to the issue of essentialism.

versions). Hull cites this example and goes on to argue as though if something as basic and important as blood is so polytypic in humans, there is no reason to suppose that any trait could exhibit the level of uniformity required for a human nature.

This kind of move is way too fast. In humans, blood groups are identified by antigens attached to the surface of red blood cells. But human blood has hundreds (at least) of characteristics in addition to surface antigens. To suggest that the case of polytypicality in surface antigens is sufficient reason to suppose that none of the characteristics of human blood are uniform across individuals is rash indeed. For starters, all human blood will contain hemoglobin. Nor is the case of antigens sufficient to justify the conclusion that none of the uniformities in human blood are also distinctive to human blood. Determining whether or not there are any such would require far more consultation with the empirical science on blood than Hull offers.

Hull is not the only thinker to wield the blood type example as a basis for rejecting human nature. In their influential introductory text to the philosophy of biology, Kim Sterelny and Paul Griffiths write:

> "So no general biological principle suggests that human moral feelings, mental abilities, or fundamental desires should be any more uniform than human blood type or eye color" (Sterelny and Griffiths 1999, 8).

What they assert is true in virtue of the fact that general biological principles are silent on the matter. No such principles suggest that the mentioned psychological traits aren't more uniform. Whether they are is an empirical question for the human sciences. But let us suppose that some human psychological characteristics are like eye color or blood type. If there really is a good analogy here, wouldn't we expect them to occur against the background of larger uniformities? That is, wouldn't we expect broader psychological

uniformities as we do with human eyes and human blood? More care should be taken in the use of physiological polymorphisms as somehow illustrative for human nature. They don't straightforwardly show what they have been taken to show.

My purpose here is not to make a positive case for, e.g., unique and universal characteristics of human blood. But the alternative inference to the absence of such is not forced on us, or so much as suggested, by an appreciation of antigen diversity. The perspective shift brought on by population thinking (a la Sober 1980) may help us better appreciate the theoretical and evolutionary importance of diversity, but that doesn't require us to deny that such diversity exists alongside uniformity. Population thinking doesn't give the kind of theoretical boost needed to license an inductive leap from instances of polymorphism to something as strong as **Distinctiveness**.

Another evolutionary consideration that ought not count in favor of **Distinctiveness** is the mere possibility of future evolution of currently distinctive human traits in other species. To illustrate, I'd like to make a somewhat conjectural claim about some distinctively species-typical traits possessed by humans which I think is likely to be true, though with some scientifically-informed refinement and increased precision.

Consider these examples of distinctive species-typical traits of humans:
Self-Control: There is some level of reflective self-control, L, such that L is not achieved by any other animal, but is achieved by virtually all statistically normal adult humans.

Frustration: For virtually all statistically normal adult humans, there is some level of self-control, perhaps different for each, that each would like to attain, but does not.

By self-control, I mean the capacity to forgo the satisfaction of some impulses and desires in order to attain the satisfaction of other, reflectively chosen, goals. Frustration is meant to express the fact that even the most disciplined of humans sometimes falters in some way. And that even the most seemingly "contentedly impulsive" can come to recognize that sometimes some of their impulses and desires frustrate others that they would count as of greater importance to them.

Call the combination of the above traits Tragic Self-Control (TSC). I claim that TSC is an excellent candidate for distinctive universality in (statistically) normally developed humans from at least the dawn of behavioral modernity. Furthermore, we have no reason to think that it will be eliminated from human populations. Note, as well, that varying levels of self-control, or desired self-control, in humans does nothing to undermine my claim. That is, it doesn't so long as the variation is within the range of L and P, where P designates a level of "perfect" self control. What we cannot rule out, however, is the evolution of TSC in other lineages in the future. But that possibility need not threaten the project of refining TSC and figuring out whether there is a version of it that is indeed true of humans, and humans alone. If true, TSC is a distinctive species-typical trait of humans now, and for the foreseeable future.

While perhaps I am wrong about TSC in particular, above I suggested that it strains credulity to suppose that amidst all the many unique features found in humans at all different levels of their biology, there is not one that is virtually universal. It may be responded that even if there are such traits, they are not likely to be the sorts of traits people have searched for in the quest for human nature. This kind of response pushes in the direction of **Propriety**, the claim that some kinds of traits are proper candidates for

human nature and some are not. Indeed, Hull explicitly denies that there is good reason to believe that there are universals unique to humans that aren't "vacuous" (1986, 4).

The Propriety of Traits

Propriety asserts that not all human traits are candidates for human nature. This implies that there are some human traits such that even if they are universal, they would not count as constituting part of human nature. Clearly a claim like this requires some conception of what human nature could be in principle. If not just any trait can be a part of human nature, then there must be some kind of substantial conception that rules some traits in and some out. It should be obvious, though, that evolutionary theory and population thinking are silent on matters of in principle conceptions of human nature. They say nothing about whether a trait would count as part of human nature if it were universal. Here it seems like the force of population thinking/evolution has genuinely given out. There are arguments to be had, at least partly philosophical in character, about substantive conceptions of human nature and the traits they propose as candidates for its makeup. Then there is empirical investigation to be done about how close to universal such traits come.[33] Besides ruling out essentialist conceptions, as already acknowledged, general principles of evolutionary theory can make no difference to the outcome of this process.

Perhaps I've misjudged the direction of the dialectic at this point, however. Hull acknowledged that one could mean something by "human nature" that is quite anodyne

---

33 I do not mean to imply that universality is sufficient (or even necessary) for inclusion in human nature. But it is a good first pass for determining which traits deserve further examination for inclusion. There ought to be really good reason to rule out universal traits as potential candidates prior to inquiry into whether or not they meet further empirical criteria. I'm arguing that none has been offered.

in comparison to such substantive conceptions (1986, 9). He was after bigger game, however: the kind of "human nature" that has been invoked by theorists and philosophers down the ages. In his eyes, it was the traditional invokers of human nature who have a vision of what traits would count as part of human nature. And perhaps he thought those traits were likely to be less than uniform throughout human populations. Unfortunately, he didn't cite a single such theorist of human nature or articulate for us even broad contours of their conceptions. Given the number and diversity of such theorists, he ought to have. It leaves us little to work with in evaluating his target.

Still, in his paper it becomes clear that Hull's target is not just human nature per se, but some of the purposes for which thinkers have employed it, especially odious ones. I will address that aspect of his argument in detail in the next part of the paper. For now, let me say that, certainly, it would be nice to undercut all nefarious invocations of human nature in one fell swoop. Showing that there is no such thing would accomplish this aim. But things just aren't so simple. Human nature qua theoretical notion deserves a theoretical hearing like any other, regardless of social purposes it may be employed for. This means, among other things, allowing ourselves the potential to be surprised in the course of investigation by the content and explanatory reach of human nature. But allowing ourselves that option requires us not to be beholden to thinkers of the past or to current judgments of intrinsic interest when it comes to the candidacy of particular traits.[34] Let me be more specific.

Do humans have violent motivations or tendencies? Human violence has been considered a candidate trait for human nature if any has. Let us suppose it turned out

34 This is so even if we also seek continuity with the broad contours of historical debates to justify continued employment of the term "human nature".

that human motivation is polytypic with respect to violence or aggression. Some percentage of humans inclines toward violently aggressive behavior and some percentage does not. So there is no uniform human nature with respect to violent behavior. That game theoretic models show that it is possible for populations to evolve stable balances between "Hawks" and "Doves" is commonly appealed to by deniers of uniform human nature (e.g., Sternly and Griffiths [1999] and Buller [2005]). Now suppose that something more or less the same turns out to be true with respect to each motivational trait that has been the subject of controversy and speculation in historical discussion (altruism, social praise and recognition, etc.). Shall we conclude then that there is no human nature with respect to motivation in general?

We should do so only if it turns out that human populations are polytypical with respect to every human motivation, not just the ones that tend to exercise the most controversy. I'll assume that there has been relatively little debate over the extent to which humans are motivated to pursue the basic means of survival. It comes as close to a universal motivation as any we are likely to find. There is a way this most fundamental of motivations tends to be realized in humans, how it interacts with their other psychological features, and how it gets expressed in behavior. It may not be as sexy a topic as altruism or aggression. But the character of the drive for survival is part of the contours of human nature if anything is. Lack of historical controversy or current scientific attention, if such there be, does not change this. Nor does whether or not the drive for survival has been exploited for fiendish purposes.[35]

---

35 I may be betraying a lack of credulity for Prinz's (2012) extreme position that would reject appeals to human nature even here. But his considerations are of a different sort than the ones under consideration. I admit that a proper response to his position would require a different mode of attack.

Any substantive conception of human nature will likely have some constraints on what kinds of traits may be considered candidates for inclusion. That is unavoidable. But the principles determining trait candidacy need to be made explicit so we can deliberate about them. None have been proposed as direct consequences of evolutionary theory and it is hard to see how any could. Bald assertions of vacuity are not helpful. But we also shouldn't be held captive by historical precedent in determining trait candidacy. On the contrary, we should be open to the idea that human nature consists of many traits that have never exercised controversy.

In addressing the more empirically based argument against human nature, I have focused solely on **Diversity** and variations that might capture different motivations behind it. I have argued that none of them are defensible, which is sufficient to defeat the argument. I have not addressed **Uniformity** and won't argue against it here. But it is worth saying that I don't think it is unassailable. That is, there may be reason to include a polymorphic trait within the domain of human nature. But I will not attempt to argue for that now. My primary purpose has been to show that insufficient reason has been given to expect that polymorphisms are pervasive in human psychology.

This concludes the first part of the paper. If my argument are successful, what is sound in the evolutionary arguments against human nature do nothing to undermine quite substantial conceptions of human nature. One, the behaviorally-modern (IBM) conception, remains viable. It can arguably satisfy two potentially conflicting desiderata: being recognizable to historically prominent proposals for a science of human nature; and being informed and refined by the actual evolutionary history of humans. It should

be clear that potential alternatives remain viable as well. It has not been part of my project to defend the behaviorally modern conception as the correct one.

We should also now be in a position to better judge whether conclusions that have been thought to follow from the idea that there is no such thing as human nature really do. I turn to that evaluative task next.

Turning Down the Bark-Volume

As acknowledged in the first part of this paper, there are conceptions of human nature for which it would be correct to say: there is no such thing. This surely includes some influential ones from history (e.g., Aristotle's?). So it might be thought that important lessons may be drawn from rejecting them. In contrast, I think there are important lessons to be learned from showing what tempting ideas do not follow. I begin by showing what little consequence there is for one of the most important debates generally associated with "human nature"—the nativism vs. empiricism debate. Then I critique specific conclusions that have been drawn concerning morality and politics, the human sciences, and kinship between philosophical movements.

Nativism vs. Empiricism

In *Beyond Human Nature: How Culture and Experience Shape Our Lives*, Jesse Prinz argues for the slogan that: "by nature, we transcend nature" (368). He intends to breath life into an empiricist view of human psychology which he thinks has been unjustifiably abandoned by large swaths of the human sciences. Does the thesis that there is no human nature, as considered here, lend any support to Prinz's project?

It should be clear now that it does not. The theoretical arguments are silent on psychological mechanisms in any species. And, as far as the principles of natural selection are concerned, any and all human psychological variation could be variation in nativist traits (however that gets spelled out). So even if the more empirically oriented arguments were successful, they would not necessarily support the empiricist point of view. Still, it is instructive to consider in more depth how evolutionary arguments bear on the nativist-empiricist debate. What matters more to that debate than the need for variation in natural selection is the gradualistic and conservative (or "cost-effective") nature of Earthly evolution.

Advances in the human and related sciences have significantly refined the terms of the nativism-empiricism debate since, say, the time of the Leibniz-Locke controversy. We cannot do justice to all the details here. What we can do, which will be important for our purposes, is see how it is possible to understand nativist and empiricist positions as ends of a spectrum. Accordingly, it is possible for different creatures' psychologies to be more or less empiricist (nativist) in an illuminating way, rather than as wholly one or the other.

The following way of characterizing the debate is especially influenced by Samuels (2002) and Margolis and Laurence (2013).[36] On this way of construing the distinction, the key issue on which nativists and empiricists differ concerns fundamental psychological mechanisms. This notion depends on a cognitive science-informed way of drawing a line between psychological and purely biological mechanisms. The fundamental psychological ones, then, are those whose operations are used to explain

[36] Prinz (2012) understands much of the debate in a very similar way (see esp. 8-14).

behavior and non-fundamental psychological processes, but are in turn to be explained only by non-psychological (i.e., biological) mechanisms. They are explanantia for psychological and behavioral explananda; explananda for biological explanantia.

Nativists and empiricists differ over the character of fundamental psychological mechanisms in two primary respects: domain specificity and developmental plasticity. Nativist mechanisms are highly domain specific. That is, they are dedicated to particular kinds of cognitive tasks, e.g., spatial navigation, numerical cognition, social cognition, object recognition, etc. Empiricist mechanisms are highly domain general. They may perform a variety of cognitive functions. Hence, nativists tend to posit more fundamental psychological mechanisms than empiricists. That is, they understand basic human psychology as more structured and richly specified.

In terms of development, nativists posit fundamental mechanisms that are more highly invariant across a specified range of developmental environments (e.g., those that can sustain full development). The empiricist end of the spectrum says that the fundamental mechanisms are much more variable over the same set of environmental conditions. So the extreme nativist posits many, domain-specific psychological mechanisms that are tightly constrained in how they develop. And the extreme empiricist posits one, or a few, domain-general psychological mechanisms that are highly developmentally sensitive to many features of their physical and cultural environments. Both extremes may be understood as anchoring two ends of a spectrum allowing for middle positions with respect to the number of domain specific mechanisms and the degree of their flexibility.

From the point of view of the arguments considered in the first part of the paper, even this updated version of the controversy may seem misguided. For, it might be thought, evolutionary theory should lead us to expect there to be variation along the scale in a way that falsifies the idea that human psychology is uniformly nativistic or empiricist. But there are a number of ways in which there might be individual variation. There could be continuous variation across the scale that is either uniform or modal (Figure 1: Panmorphic 1 & 2).[37] There could be polymorphic variation around a few nodes (Figure 1: Polymorphic). Or there could be variation clustered around one node (Figure 1: Monomorphic). Over time, the node could move in one direction or another. All are consistent with the general requirements of evolution by natural selection. So the nativist-empiricist controversy should only seem misguided if good reason is given to expect that the monomorphic option is not likely, or that the polymorphic option is more likely. But none such arguments have been offered.

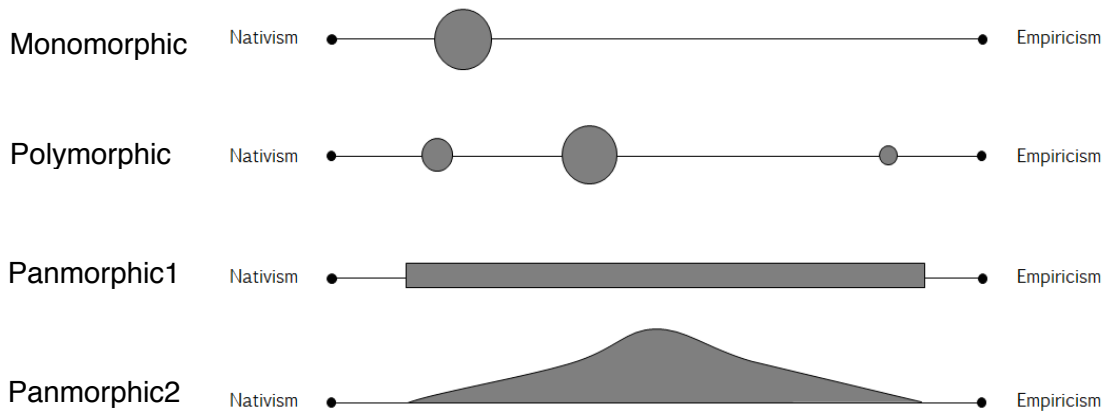

Figure 1: Possible distributions of individual psychologies

In fact, comparative ethology and cognition give us good reason to think that wherever it happens to be on the scale, human psychologies are likely to be more

[37] Thanks to Colleen Sullivan for help with the figure.

monomorphic than not. Perhaps surprisingly, this is because, from a sufficiently general but well-motivated standpoint, a large proportion of all species capable of perception and learning are of the same "psychmorph". And this thanks to the fact that there are general principles that characterize Earthly perceptual and learning systems (Shettleworth 2013, 1998; Hauser 2001; Heyes 2012). Examples include principles of habituation, constancies, associative learning, and memory (ibid). And the commonalities multiply as cognition becomes more complex. Humans share features of "core cognition" regarding object recognition and quantitative and spatial cognition (among others) with many species they are distantly related to and more such features with ones they are closely related to (Carey 2009).

Of course, we do not want to go too far, and deny that there are not very important cognitive differences between species. But when we are asking a very general question like: do biological psychologies tend to be more nativist or empiricist?, then such differences are less important. If the question is whether species tend to have specialized, dedicated cognitive mechanisms with highly canalized development, then it is not as significant whether a species has a particular one that others don't. If the question is whether perceptual systems tend to be geared into learning mechanisms governed by common associative principles, it far less relevant whether a given sensory modality in a species is sensitive to a wider spectrum of stimuli than in others. Those kinds of variations are very important for other kinds of questions, but not to the nativist-empiricist controversy. As the evidence stands, animal psychologies tend to be a combination of generalized associative learning mechanisms and specialized cognitive

mechanisms (Shettleworth 2013; Heyes 2012). They are somewhere in the middle of our nativist/empiricist scale.

When it comes to humans, it is clear that they are able to learn a lot more kinds of things and have much more flexible behavior than other creatures. The best explanation may require shifting the human species closer to the nativist or empiricist ends of the scale. I won't take a stand on that issue here. What we shouldn't expect to see is that the uniquely human capacity for learning and flexibility is the result of wide, ongoing, and continuous variation toward both ends of the scale. Or the evolution of widely divergent polymorphs on the scale. Given what we know about the conservation of brain and cognitive architecture over evolutionary time, and the generally gradual pace of it, variation of this degree is highly unlikely (see Allen 2009). Appreciation for variation inspired by principles of evolution by natural selection do not, cannot, render the nativism-empiricism debate obsolete. Rather, general facts about how it tends to work in the biological material we are familiar with can reinforce specific answers to the debate.

So one lesson to learn here is that it is not simply the principles of evolutionary theory as they could be featured in systems of very different materials, organization, and so on. Rather, we need to think about principles of evolution of stuff around here. Radical changes in the basic organizations of most things is very expensive. The mechanisms of evolution, whether natural selection or drift, will be more likely to tinker than to overhaul. Insofar as the traversal of large distances on the nativist-empiricist scale requires psychological overhaul rather than tinkering, we should not expect much variation. The further and further we zoom in on a species, the more variation we may

discover. But the features of fundamental psychology are more likely to vary as the basic organizational features of Mammalian body plans do; not the way Mammalian body size does.

I'll conclude this section by returning to Prinz's project. The kinds of arguments against human nature I consider in this paper would undermine his project just as much as they would any nativist. For they are arguments targeted at the uniformity of human psychology and not its broad contours. There can be no reinforcement between empiricist denials of human nature and evolutionary denials of human nature.

Morality and Politics

In "On Human Nature", David Hull (1986) notes the following intellectual function that human nature can have:"But why is it so important for the human species to have a nature? One likely answer is to provide a foundation for ethics and morals" (11). As he thinks that human regularities are certain to change over time, he says: "I, for one, would be extremely uneasy to base something as important as human rights on such temporary contingencies" (4). He also expresses befuddlement that people even think that things such as rights need to be based on commonalities at all. Still, Hull appeals to evolutionary-based denials of human nature to undercut immoral treatment of particular groups of humans. I will quote his thoughts on the topic at length (from an earlier paper):

> "On the historical entity interpretation [of species], retarded people are just as much instances of Homo sapiens as are their brighter congeners. The same can be said for women, blacks, homosexuals, and human fetuses. Some people may be incapable of speaking or understanding a genuine language; perhaps bees can. It makes no difference. Bees and people remain biologically distinct species. On other, nonbiological interpretations of the human species, problems arise (and have arisen) with all of the groups mentioned. Possibly women and blacks are human beings but do not "participate fully" in human

> nature. Homosexuals, retardates, and fetuses are somehow less than human. And if bees use language, then it seems we run the danger of considering them human. The biological interpretation has much to say in its favor, even from the humanistic point of view" (Hull 1978, 360).[38]

Interpreting his remarks, we can say that Hull proposes to draw the following lessons for ethics and politics:

Instability There is not enough stability in human features to play a part in determining morality and ethics.

No Normative Classification Biological classification of the human species removes any biological reason to classify any member of Homo sapiens as not-properly-human in any sense.

Let me begin with *No Normative Classification*. It is certainly true that current biology removes any presumed biological basis for classifying large groups of humans in hierarchical ways with moral and political significance. And there is certainly something very attractive about being able to wipe away any "natural" or biological reason to discriminate against historically oppressed groups in one elegant stroke. Furthermore, it is a stroke offering insurance against future inventions of groups to discriminate against as well. Still, things aren't quite so simple. Ironically, the arguments for polymorphisms may turn into a double-edged sword here. Suppose different "psychmorphs" were discovered in human populations. Sad experience suggests we can expect those to come along who are motivated to treat one of the psychmorphs in some discriminatory manner, and who appeal to evolutionary theory as a basis for treating them as a naturally distinct subspecies or "race" of human. Saying that the oppressors could only

38 The 1986 paper mentions the same groups and expands on the point.

do this based on a misunderstanding of evolution is not to the point. When has oppression ever been based on an accurate understanding of biology?

Furthermore, holding that every bit of the *H. sapiens* lineage automatically gets the moral status usually associated with "human" may have problematic consequences for any being outside that lineage. If that becomes the means for determining moral status, it becomes irrelevant what psychological characteristics beings outside the *H. sapiens* lineage possess. The fact that all members of a lineage get grouped together in a certain way by biology does not entail that they should all be similarly grouped for moral or political purposes. To think otherwise is to exaggerate the reach of biological classification once it has been re-conceived in the way Hull envisions. If evolutionary theory allows for different types within a lineage at all, it cannot by itself prevent us from finding morally relevant differences between the types. It certainly can't stop those who willfully misunderstand or distort biology for depraved purposes from finding something to latch on to.

The upshot, I think, is that there is no shortcut around substantive moral theorizing for telling us what natural traits, differences, etc. are morally relevant and what are not. We should not expect a silver bullet from biology to do away with attempts to ground odious ethics and politics in biology.

Next, it should be clear now, from the previous arguments, that no one has come anywhere near establishing something like *Instability*, the thesis that there is not enough stability in human biology for it to play some role in determining the character of our ethics or projects of establishing just and flourishing societies. Perhaps more importantly, Hull does not give adequate reason to suppose that there is no positive way

that general facts about humans can contribute to morality. Suppose, as seems plausible, that some suitable ought implies can principle applies to individual human morality, and political governance as well. So, for instance, individuals can not be morally obligated to do things they cannot do. It is also not a good idea to undertake social programs or form political institutions that are doomed to failure. There are surely many things that humans cannot do, nor will be able to given any likely course of evolution. Further, it may be that there are many such things we are mistaken about or unaware of, some of which may conflict with extant moral theories. If so, "the rejection of human nature" does not rule out inquiry into what humans are generally like from contributing to our moral and political theories.

Bottom Line: The denial of human nature cannot straightforwardly undercut nefarious appeals to human nature. In fact, it even leaves open the possibility of looking for biological differences between groups of humans to base moral theories on. Further, it cannot and should not rule out some potentially desirable consequences of using human biology and psychology to inform our moral and political philosophies.

Human Sciences

While many philosophers, biologists, psychologists, anthropologists, and other scientists have thought that evolutionary theory has profound implications for how we approach the human sciences, Hull and Buller think that most such enthusiasm is misdirected. They think the discovered lack of human nature ought to reorient the human sciences in very different ways. Psychology should cease the search for universals (Buller 2005); nor should anthropology search for cultural universals (Hull 1986). And they suggest ways of reorienting both disciplines away from sciences like

physics in order to become more "historical" sciences. We'll look at each to see how they overstate their case.

Buller on Psychology

After criticizing certain prominent work in evolutionary psychology, Buller (2005) reaffirms the idea that psychology should be based in evolutionary theory, properly conceived, and outlines how it would change when set aright. To achieve the proper direction, he claims that "…a truly evolutionary psychology should abandon the quest for human nature and with it any attempt to discover universal laws of human psychology" which latter he equates with "…the quest to be a science in the model of physics or chemistry" (419, 457).

Buller goes on to accuse the search for universals of a kind of "temporal provincialism", even suggesting that it retains holdovers from theological creationism. He claims: "…any psychological universals we might happen to discover in the present would not characterize Homo sapiens per se, but would only characterize our species at this particular moment in evolutionary time" (477). In light of this situation, he suggests that psychology wake up to the realization that "evolutionary theory is purely a process theory" addressed at generational change rather than "finished products" (478-9). Although he does not go into specifics, it is clear that, somehow, psychology needs to be more historically oriented, becoming a science of process and change.

There are a number of problems with Buller's arguments. First, it is bizarre to suggest that evolutionary theory, as opposed to physics and chemistry, is directed at natural processes and change. Similarly with the implication that psychology does not already itself deal with processes and change (consider developmental psychology). It is

also misleading to suggest that evolutionary theory is purely aimed at processes. As elsewhere in nature, evolutionary processes have certain kinds of outcomes or "products", and not other kinds. Biology is just as concerned with the biological outcomes of evolutionary processes as with the processes themselves.

With the above in mind, it is also misleading to suggest that biology has become a completely "historical" science because of evolution. From microbiology to ecology, biology employs categories that are not purely historical in nature (e.g., cell, sexual dimorphism, autotroph, predator) and discovers principles of them that are invariant across many kinds of evolutionary change (e.g., ratios of biomass of autotrophs to heterotrophs in ecosystems). Natural selection itself often drives traits to fixation in populations and maintains them at fixation. Hence, far from denying all forms of biological "stasis", in many cases evolutionary theory both predicts and explains it.

Next, numerous biological traits and principles operate over a large number of species (note that medical and nutritional principles tested on animals such as mice often generalize to humans). The same goes for psychology, undercutting Buller's claims that any universal is destined to be temporary. Principles of both associative learning and faculties for "core cognition" are operative throughout humans and members of other species, many of which are only distantly related to humans (Shettleworth, Heyes, Hauser, op. cit.).

It is highly likely that some currently typical features of human psychology are likely to become (statistically) untypical in future populations. That fact alone does not even make it probable that all will. Given what we have learned from comparative psychology, many psychological features of humans are likely to remain typical in

human populations so long as they exist. And if psychology can discover psychological traits and principles that operate universally across species, there is no reason that it couldn't discover some that are both universal and unique within a species. Buller has failed to identify a way in which psychology needs to be radically reoriented to accommodate the advances of evolutionary theory. It may recommend greater caution and higher standards for claims of universality, as well as greater precision in the scope of the "universality" being claimed. But it gives us no reason to give up on psychological universals tout court.

Hull on Anthropology

Hull claimed a near universal tendency on the part of anthropologists to search for cultural universals, and chided them for it. He believed that biology held the antidote. He said: "But if genetic variability characterizes species, even though everyone is absolutely certain that it does not, then possibly a similar variability characterizes cultures, even though the parallel conviction about cultures is, if anything, stronger" (5). He explicitly developed the analogy further:

> "My argument is analogical. Both population geneticists and anthropologists have been strongly predisposed to discount variability. Genetics is sufficiently well developed that geneticists have been forced to acknowledge how variable both genes and traits are, both within species and between them. The social sciences are not so well developed. Hence, it is easier for them to hold fast to their metaphysical preferences" (ibid).

As an assessment of the field of anthropology, Hull is profoundly one-sided. The most historically influential tradition in American anthropology, starting with Franz Boas, was characterized by its emphasis on cultural diversity (Degler 1991). However, many of the stronger claims of variability and lack of commonality have not stood up to the empirical "development" of the field (see e.g., Brown 1991; Pinker 2003). But what is

hard to understand is why Hull thinks that whatever variability there is in human genetics is relevant as a guide to variability in cultures.

Perhaps there is some causal route from the variability of genes that would lead us to expect variability among cultures.[39] For there may be just as much variability in genes involved in psychological development as in anything else. If this variability led to variability in psychologies across cultures, and cultures reflect the psychologies of their members, then perhaps we should expect to see genetic variation expressed in cultural variation.

There are many obstacles for the above argument to overcome. But even if it turned out to be accurate, so that we should expect to find cross-cultural variability, it (once again) does not rule out finding universals amidst the diversity. It could be that any sustainable culture requires some broad organizational principles and practices to maintain itself and to transmit its knowledge from generation to generation. Further, as argued above, there is no reason to think that psychological universals themselves have been ruled out by Hull's considerations. In principle, anthropology could safely proceed in simultaneously searching for cultural universals while documenting the array of discovered differences.

Bottom Line: Neither Buller nor Hull give any good reason to suppose that human sciences, such as psychology and anthropology, cannot follow physics, chemistry, and even biology in employing non-historical categories and discovering broad generalizations that make use of them.

Philosophical Kinship

---

39 Hull denies that he is asserting this. But it is not fully clear just why he thinks it is relevant and this is an option at least worth exploring.

In his recent introduction to the philosophy of biology, Peter Godfrey-Smith (2014) tries to connect the evolutionary and existentialist approaches to human nature. He identifies their commonality as a "post hoc" approach. Specifically, human nature talk is always a response to already established human phenomena that are, in some sense, changeable and open-ended. From the existentialist side, he quotes Sartre:

> "What do we mean by saying that existence precedes essence? We mean that man first of all exists, encounters himself, surges up in the world—and defines himself afterwards" (Sartre 1945; quoted in Godfrey-Smith 2014, 143).

Godfrey-Smith interprets Sartre as holding that "there is no human nature that does or should constrain the actual facts of human behavior and choice. Humans are what they make of themselves" (142). He then immediately connects this with how he had earlier characterized "evolved natures" as post hoc, suggesting endorsement. The earlier characterization is as follows:

> "Once evolution in a lineage has actually taken a particular path for a while, we can talk about an "evolved nature" that has been established in that lineage, though much of it will not be universal, even at a time. As evolution is open-ended, this talk about our nature has a post hoc character" (142).

There is a simple, literal interpretation of Sartre's and Godfrey-Smith's claims that is true, but uninteresting. All talk of natures, whether human, biological, or otherwise, occurs after humans have begun to exist, evolved, and "surged" up in the world. Before them, there was no one else to do it. The important question, though, is one of whether that talk latches on to genuine discovery of non-arbitrary groupings of phenomena. The fact that evolution can change currently universal features of kinds of organisms is what makes zeroing in on all current universals appear to have an element of arbitrariness and, hence, be post hoc. But this open-endedness of evolution occurs throughout the

biological realm. And the existentialists intended to distinguish humans from the rest of life by means of their "self-made" characteristics.

Godfrey-Smith recognizes the "universality" of the open-ended nature of evolution, and so suggests that "The capacities for learning and cultural transmission seen in humans give this evolutionary openness an extra dimension" (142). It is doubtful whether this really connects with any existentialist viewpoint, particularly Sartre's. For Sartre, it wasn't any peculiar capacity for learning, but a particularly powerful kind of freedom to impose meaning that gave humans their self-making capacity (Sartre 1945). And many of the human tendencies involved in cultural transmission would be understood by Sartre as expressions of "bad faith", or tendencies of humans to disown their freedom (Sartre 1943). Furthermore, while the main motivation behind evolutionary denials of human nature stems from its emphasis on diversity and a lack of unity in the H. sapiens lineage, Sartre thought all humans were unified by their powers of "nature-transcendence" (1945).

The tendency to think of human flexibility and control over their lives and environments as a threat to traditional notions of human nature does not actually begin with the existentialists. Kant and Marx are two towering figures of the modern era who held views of this kind.[40] Anthropologists, particularly in the American tradition, also argued for similar views (Degler 1991). Advances in biology and the human sciences tell

40 For Kant, see his 1798, 1786, and Wood's 2007 interpretation. For Marx, see his 1845, 1846. See Solomon 1972 for an explanation of Kant's profound historical influence on the rise of existentialism.

us that the extent of human plasticity, flexibility, and self-control has almost certainly been exaggerated in these various traditions.[41]

Still, it should not be denied that humans have more, or a higher-degree of, "self-making" capacity than other creatures. Spelling out just what this comes to in much more clear and precise terms remains as an important task for philosophy and the human sciences. The way I see it, there remains a need to adequately characterize the extent and limits of human self-making, at both individual and collective levels, as well as a need to then explain what gets so characterized. The importance of the conceptual and theoretical reorientation accomplished by evolutionary thinking to these tasks resides more in the theoretical possibilities it brings into view for these tasks, rather than anything it tells us about human beings in particular. Any convergence between evolution-based and other philosophical denials of human nature is purely nominal, and thinly so, at best.

Bottom Line: The rationale behind evolutionary denials of human nature bears no significant similarity to other philosophical movements that call into question the very notion of human nature. The former denial is based in facts that humans have had no control over, while the latter sorts of denials are based in an appreciation for what control humans do have, or at least appear to have.

Conclusion

---

41 Pinker 2003 remains a useful documentation of this, even if somewhat exaggerated. Banaji and Greenwald 2013, Kahneman 2011, and Wilson 2002 are excellent summaries of the psychological evidence indicating just how much inflexibility, opacity, and bias pervades the human mind. Even the bastion of Kantian and Sartrean freedom is riddled with error, illusion, and inability to command a clear view of its own limitations.

The strength of the evolutionary case against human nature has been greatly exaggerated. To delegitimize any talk of human nature in the biological and human sciences requires much stronger premises than deniers of human nature have either recognized or adequately argued for. Alternatively, the arguments that may rule out certain, misguided conceptions of human nature from the past have relatively weak consequences. They don't have the implications for currently live issues that they have been thought to have. Furthermore, answers to important questions relevant to the issue have been assumed rather than treated as topics for investigation in their own right. One such issue is the actual extent of variation in human populations and how to adequately characterize its co-existence with uniformity in the same populations. This may seem like a straightforwardly empirical question. It certainly has a large empirical component to it. But the fact that it continually arises in debates across the human sciences suggests that philosophers may have an opportunity to contribute to making progress on the issue. If the debate over human nature in philosophy of biology helps bring focus and clarity to this issue, then that will be one fruitful consequence of its legacy.

SHIFTING ATTENTION ON TRANSFORMATIVE CHOICE: THE ROLE OF WORTHWHILENESS IN MAKING THE LEAP[42]

Introduction

In recent work, L. A. Paul (2015; 2014) has argued that some important decisions, such as whether to become a parent, cannot be made rationally when approached in a common way. We can neither rationally choose for or against these particular actions. The reason is that we have no rational basis for evaluating certain ways we will be affected by the consequences of our choice. Rational basis is missing when our choices involve undergoing dramatically new kinds of experiences and those experiences change our preferences and values in ways we cannot anticipate. These experiences are both epistemically and personally transforming. In Paul's words:

> "Any epistemically transformative experience that changes the self enough to generate a deep phenomenological transformation creates significant trouble for the hope that we could use our ordinary subjective perspective to make rational decisions about major life events (2015, 22)".[43]

The problem is that transformative experiences create values that are inaccessible to the inexperienced, values that the inexperienced person needs in order to make a decision about whether to become experienced or not.

There are two general ways of responding to Paul's problem. One is to try to show that the purportedly inaccessible values are somewhat accessible after all. This strategy has been pursued from different angles in some of the forthcoming commentary on Paul's work (Barnes, forthcoming; Dougherty, Horowitz, Sliwa, forthcoming; Pettigrew,

42 I am grateful to Ned Hall, Bernhard Nickel, L. A. Paul, Susanna Siegel, and Susanna Rinard for very helpful discussion on this paper and its topics.

43 For Paul, "deep phenomenological transformation" implies personal transformation.

forthcoming). A different strategy grants, at least provisionally, the inaccessibility of the values and proposes alternative bases for making the decision. This is Paul's own strategy, and the one I engage with her on in this paper. Her proposed decision base is the value we place on discovery of inaccessible values. I raise two problems for this proposal, one of which is a major challenge to any alternative proposal. I propose a way to meet the challenge through including in our decision base the values we place on the inherent worthwhileness of the transformed identities, activities, and relationships. First, I turn to an explanation of Paul's problem and how to understand its force.

Paul on Transformative Experience

What, exactly, are these inaccessible values? They are the values we attach to the phenomenological features of brand new experiences. Their inaccessibility stems from the fact that the phenomenology of a type of experience is epistemically blocked from all but subjects having the experience. If you have never eaten insects before (a common food in many parts of the world), you're first bite into a grasshopper will reveal to you what it is like to taste grasshoppers. Experiences of new tastes, sights, sounds, feelings, emotions, etc., are 'revelatory' in that they reveal to us the intrinsic character of new phenomenal states (2014, 13). They have further epistemically transformative properties in that they also endow us with new capacities of imagination and recognition (2014, 10).

As Paul notes, we value our experiences, especially new ones, and they "carry weight in our decision making" (2014, 11). While such values will include hedonistic ones of pleasure and pain, they are not limited to them. Paul argues that additional

values we attach to the intrinsic sensory, affective, and cognitive character of phenomenal states are "primitive"--not reducible to anything else, including hedonistic values (2014, 11-12, 93). She says: "Because I take such values … to capture the rich, complex nature of lived experiences resulting from our sensory as well as our nonsensory cognitive phenomenology, I will describe them as *subjective values*" (12, original emphasis).[44] This gives a good sense for the rich approach Paul takes to our experiential lives and how we value them. But I wish to avoid potential confusion that may result from labeling these values 'subjective'. All values in decision-theoretic contexts, not just those assigned to experiences, are "subjective" in the sense that they are assigned from the point of view of the decision making subject. So I prefer to use "experiential values", but will mean by it exactly what Paul does in her use of "subjective values".

Let me be clear that what is 'inaccessible' to a decision making subject in the first instance are the phenomenal states 'on the other side' of the epistemically transformative experience she is considering. Paul's crucial contention is that there is a rational constraint on the assignment of values to those experiential states. The constraint is that we have to be subjectively acquainted with revelatory states in order to be justified in forming preferences concerning them, or to assign specific decision

44 Paul places significant emphasis on cognitive phenomenology and takes for granted its non-reducibility to any kind of sensory phenomenology (2014, 12). She does not, however, address the fact that this is a very controversial assumption in philosophy of mind (though she does cite Bayne and Montague [2011] where the controversy is on full display). It would be interesting to know what she thinks about how her project would be affected by the outcome of this controversy. However, although I can't argue it here, I think her inaccessibility problem can be generated without such reliance on cognitive phenomenology.

making weights to them.[45] This is the primary sense in which experiential values are inaccessible for Paul: even if we can denote and talk about them ourselves (perhaps because of testimony), we cannot rationally assign them for ourselves without the proper, first-personal epistemic acquaintance. There is fairly strong intuitive backing for this claim. If you were to ask me which I prefer, vegemite (a spread) or durian (a fruit) [examples used by Paul], I would respond that I cannot say for I have never tried either. And it would be unreasonable for you to demand of me that I go ahead and form a preference anyway.

Despite the inaccessibility of the taste of durian, it doesn't seem irrational for me to decide to try it. I want to find out what it tastes like, and past experience with hundreds of new tastes gives me considerable confidence that it will not be so bad as to egregiously displease or harm me (see Paul 2014, 15 and 37-39). Things change, though, when I must decide whether to undergo a transformative experience that is much more foreign and has much greater consequences than tasting a new fruit. Decisions like whether or not to parent can potentially change the very basis on which I make value judgments, reconfiguring my whole preference structure. They may deeply affect my sense of self, the very character of what it is like to be me. When experiences have this outcome, they are personally transformative in addition to being epistemically so (2014, 16; 2015, 8). Becoming a parent; going to college; living amidst an unfamiliar culture; religious conversion; becoming clinically depressed: these are illustrative examples of personally transformative experiences.

---

45 I find the best statement of this to be the whole last paragraph of page 13 of her 2015. Of course, it is an ongoing theme in both the paper and the book.

The first important thing to note about how personally transformative experiences differ from minimally epistemic ones like trying new tastes is how much higher the stakes are. With a new taste, we know that the potential effects are acceptably limited or containable. With a major life decision like parenting, however, we know that the effects, both objective and subjective, can be far-reaching. With so much at stake, we want more information to be able to make the right decision. But unfortunately there is even less; or, rather, there is more ignorance. There is more 'breadth' to the ignorance: the novel kinds and aspects of phenomenal experience for a new parent greatly outnumber those of a new taste.

There is also a special kind of 'depth' to the ignorance. Whereas my current psychology is "primed and ready" for the experience of tasting durian, it is not so ready to hear the cries and coos of an infant as communicative signals of the current well-being of my child. My psychology is as yet unable to adequately process such an experience and evaluate it properly. Until I have undergone many hormonal and neuronal changes that will affect my senses, desires, and preferences, some of the transformed experience of parenting will be 'doubly' inaccessible to me. If I had an 'experience-scope' to peer into the mind of a parent, some of it may be like watching a Japanese Noh play to the uninitiated--I would flounder to even make sense of what was happening. And so without the transformed psychological apparatus that comes with being a parent, there is a sense in which the values attaching to parental experience are deeply indeterminate for my current psychology in a way that the value of tasting durian is not (see Paul 2014, 48-9). In the next section I will discuss in more detail what Paul thinks this means for rational decision making.

Paul on Contemporary Decision Making

Paul holds that a decision is rational only if made in accordance with what she calls realistic normative decision theory (2014, 19-20). For her this means standard Bayesian decision theory tailored in whatever way needed to be usable for beings with our cognitive capacities and limitations.[46] And, though she affirms that "maximizing expected value is the way to go" in matters of rationality, decision theory cannot be sufficient for rationality without qualification for her (2014, 114).[47] This is because she has the constraint of epistemic acquaintance on rational formation for experiential preferences that we noted above. So that constraint needs to be included in the total list of jointly necessary and sufficient conditions for rational decision making. Still, with that in mind, we can state Paul's problem of inaccessible values in decision-theoretic terms:

No Output

If the basis for a personally transformative decision consists of transformed experiential values, then decision theory yields no expected values for the decision options.

The reason for **No Output** is that the relevant experiential values are inaccessible and so we are not 'in possession' of values to plug into the decision-theoretic calculus. Without the proper inputs, we cannot get the needed outputs. A decision problem that

46 She mentions Weirich (2004) as a good source for how to approach this.

47 Note that she uses "expected value" rather than the oft used "expected utility". This is purely notational as far as I can tell and will follow her in it. At the very least it does not signal a departure with standard decision theory that need concern us.

cannot get expected value outputs from decision theory cannot be solved rationally by anyone taking decision theory to be necessary for rational decision making.[48]

The antecedent of **No Output** mentions a basis for transformative decisions. What determines what values to base a decision on? Paul cites a contemporary cultural ideal, claiming it is the reigning cultural paradigm for how to make decisions like the parenting one (2015, 12; 2014, 25). Whereas people in other times and places relied on values of economic necessity or cultural values of preserving family status, many contemporary Western societies enshrine ideals of personal fulfillment and self-realization (2015, 21; 2014, 84). She interprets these ideals as directing us to base transformative decisions on "…what it will be like for us to experience the outcome of our acts…" (2014, 25). That is, our attention to experiential values is being directed by commitment to a well-entrenched ideal of self-realization.

Another route to experiential values comes from Paul's interpretation of a natural way for us to make decisions which is by imagining different "subjective futures" that could result from our decisions (2014, see esp. 33 and 52). Paul speaks of this as cognitive modeling, wherein we "evaluate … experiential outcomes" of our possible acts by mentally simulating "…what it would be like for (us) to have each of these

48 Pettigrew (forthcoming) proposes a way to formally model the uncertainty involved that challenges No Output. At best, the model shows that it would be too quick to suggest that there are any deep, novel problems here for ideal decision theory. But the model is highly unrealistic for everyday decision making that employs a suitably real-life-agent-tailored decision-theoretic calculus. Even then, I'm not sure that it does get ideal decision theory off the hook, for it assumes a method for weighting temporally indexed utility functions that needs cashing out (see Paul's forthcoming response). In any case, if my proposal here works, it provides a realistic way of gaining traction over the uncertainty in transformative decisions that allows us to avoid the challenges of employing Pettigrew's model. Thanks to Susanna Rinard and Ned Hall for discussion here.

experiences" (26). It is important to be aware of these two sources of focus on experiential values to know where to return to when we call into question that focus.

Note that the previously mentioned antecedent contains "consists", indicating that only experiential values are part of the decision base. This will be important to how we think about addressing the problem of inaccessible values--whether to completely replace them, or just force them to cede room to other, accessible values. In this paper, I will be assuming, along with Paul, that we cannot simply ignore or replace inaccessible experiential values. The problems I raise for both Paul's and my own proposed solution depend on it. But this means that whatever additional or alternative values are incorporated, the decision-theoretic issue of how to deal with unassigned values that ought to be part of the calculus will arise. This issue will have to remain in the background for the remainder of this paper, though. We need to get other proposals all the way on to the table before we address the logistics of decision-theoretic calculation.

Two issues concerning decision making remain before we can get to potential solutions to our problem. The first is Paul's commitment to another norm for decision making. In addition to making them rationally, she thinks we should make decisions authentically. This ideal asks that we strive for 'faithful' representation of our desires, values, and preferences (2014, 105). That is, we resist or root-out self-deception and willful self-ignorance. Additionally, it asks that we retain autonomy and authority over our decision making and refuse to cede it to alien sources (2014, 130). For Paul, this has the consequence that we cannot replace our own doomed first-personal efforts at accessing the inaccessible with scientific findings (possible in the future, though not advanced enough now). She wants neither "Big Data" nor "Big Morality" to replace

"personal deliberation and authentic choice" (2014, 130). I am in broad agreement with Paul on this and my own proposal is intended to respect authenticity so conceived. As I said in the introduction, I will not be concerned with any ways of rendering the inaccessible values more epistemically accessible.[49]

Finally, Paul is only concerned with transformative choices that are not decided morally, legally, empirically, prudentially, or economically (2014, 18, 112; 2015, 1). Once all those kinds of considerations are in, there must be options left over which people can rationally disagree about. Rationality by the lights of decision theory, after all, allows different people to choose differently in the same decision problem. In the search for alternative bases to making transformative choices, we should look for general guiding ideals over which people can disagree in the particulars. Self-realization meets this test--different people will choose alternative routes to self-realization and each may be rational.

Now we should have a good grasp on Paul's problem of inaccessible values and how it threatens the rationality of personally transformative choices. The problem can be persuasively articulated both informally and in the context of standard decision theory. We should also have an awareness of important nuances in the structure of the problem to help us think about what is needed to meet the problem and evaluate whether different proposals can do so. I turn to that task in the remainder of the paper.

Paul's Solution: Discovery of Experience

49 However, I do think that there are options for reconciling a reliance on scientific evidence with the value of authenticity that Paul does not consider. But that will have to await another occasion.

Paul's response to the problem of the inaccessibility of transformed values is to propose the discovery solution. She says we can rationally approach the problem by asking ourselves how much we value revelation for its own sake. In the choice between parenting and not, do we prefer discovering what it would be like to be parents to being forever ignorant? Do we value discovery enough to override the value of what it is like to be us in our childless states? If not, then we should forego parenting. Similarly for any decision in which transformation is a live option. How much do we value discovering the results of the transformation?

Discovery has value for virtually everyone to some extent. Paul notes Aristotle's essential insight in this connection: "All men by nature desire to know. An indication of this is the delight we take in our senses; for even apart from their usefulness they are loved for themselves…" (quoted on pg. 92 of her 2014). We all know people driven by exploration and discovery. Science depends on them. "Now I can say that I've had the experience of Xing"; "I've always wanted to try Y" are phrases commonly heard for numerous and various X and Y. Discovery value also has the advantage of being one we recognize different people having to varying extents. It doesn't dictate that each person should parent, or refrain from parenting. And this is a result we want.

Approaching transformative decisions through discovery values is not simply like deciding whether to try bungee jumping. Though there may be some parallels, Paul emphasizes that the decisions cannot be reduced to sensory discovery or thrill: "If you choose to have the transformative experience, to choose rationally, you must prefer to discover whether and how your preferences will change" (2014, 118) [*mutatis mutandis* for choosing against transformation]. One must be fully aware that currently cherished

first-order preferences, perhaps ones for sensory adventure and discovery, may be lost. Other important things are at risk as well: "…the decision to have a child could be understood as a decision to discover a radically new way of living with correspondingly new preferences, whether your subjective well-being increases or not" (2014, 119). As she says, "…you choose [transformation] in order to discover who you'll become", where this could be different in quite significant respects from who you are now (ibid). One must value discovery highly indeed. But such discovery, of fundamental changes in one's sense of self, may be one of the privileges of being human.

Discovery values have further advantages from Paul's point of view. For one, they are not inaccessible. We have all had the experience of discovering something or other. Revelatory experiences made up significant parts of our lives from infancy through adolescence. If Aristotle was right, there may be some native basis for valuing discovery and so it may make little sense to question its rationality. Hence, we have rational preferences right here and now for discovery. We need not pretend to know enough to value that which we aim to discover. Nor do we need science or other testimony to tell us how much we value discovery. We are not alienated from how much we value discovery. Additionally, we may have had personally transformative experiences in the past. Anyone who has gone through puberty knows what it is like to be transformed. And while that may not have been a choice, it provided a store of transformative experience to assist us in forming values for discovering personal transformation. Basing transformative choice on mindful consideration of the value of transformation discovery thus seems to have much to recommend to it. Unfortunately, it has significant problems of its own.

Troubles with Discovery

In this section I will raise two worries for the recommendation to make transformative choice rational via the consideration of the value of discovery. I call them the guidance problem and the gravity problem. I will begin with the guidance problem and consider some possible responses to it. I conclude that none are likely to succeed. Then I turn to the gravity problem.

Guidance

Recall that the discovery solution Paul recommends requires basing transformative choices on the value of transformative discovery per se. She says it: "The idea is that if you choose revelation, you choose it for its own sake" (118). And it is required by both her view and the nature of discovery: you choose to discover epistemically inaccessible things which have inaccessible values. But the value of discovery as such cannot help you decide whether to discover X, Y, or Z, where these are mutually incompatible options. It can only help you decide between discovery and foregoing discovery. Hence, from the point of view of discovery, you should be indifferent between discovery of what it is like to become a parent and discovery of what it is like to become a celibate monk. It cannot tell you whether to discover who you would become as a lawyer or to discover a future self as a dentist. It is not that it values them in different, incomparable ways. The value of discovery simply has nothing to say. It can give you no guidance amidst the array of alternative transformations.

Possible Responses

The value of discovery can help us decide when the decision problem has been legitimately delimited like so: to discover or not discover. The problem is that it cannot guide us in situations where we must choose between alternative paths of discovery, and those are more realistically the sorts of situations we find ourselves in, e.g., choosing career paths. I want to now consider some possible responses on behalf of the discovery theorist. Two kinds of response I examine fall under the category of augmentation strategies. Such strategies seek to retain discovery as the primary motivating value, but pare down the space of alternative discoveries through additional considerations and values. The second category of strategy, a 'dogmatic' or 'dig-in-your-heels' one, insists on the rationality of 'brute' preferences for the discovery of X over the discovery of Y. In each case I find the strategy unable to deliver.[50]

Augmentation Strategy #1 Suppose you have a very strong first-order preference, along with a higher-order preference that it not be changed. You also value discovery strongly, but not enough to change the other first-order preference. In other words, you value discovery enough to be willing to change some first-order preferences, but not all. Then you can rule out all the discovery options for which there is a significant chance your highly valued first-order preference would either be changed or continually frustrated.

Let's start with my example of parenthood versus monkhood. A response to it following the above strategy would be to say that preferences concerning sexual activity are both rational and strong enough to rule out discovery of experiential values resulting from changes in them. This kind of response may work against this particular example.

---

50 I'm grateful to Susanna Siegel for discussion that helped motivate my formulation of the possible responses here.

But it cannot do much broader work. The space of potentially transformative choices is not significantly restricted by fixing and respecting the strength of basic sexual preferences (e.g., for some sex over none). Choice of career is one domain that is not plausibly so restricted. With that in mind, it is also worth considering how Paul's argument for inaccessible values more directly confronts the response.

If you think that you know you would not sufficiently value the experience of living the life of a celibate monk, you are not yet convinced by Paul's inaccessibility arguments. The whole point is that personal transformation changes you in ways that you cannot adequately anticipate and so reject out of hand. This is no less true for becoming a parent than for becoming a monk. Choosing to remain childless requires choosing to reject experiential values for which you have no basis of rejection (or acceptance). It is exactly the same for deciding to reject monkhood. If you became a monk, your first-order preferences and values would change in such a way as to change how you value certain experiences. As a monk, you may come to enjoy cognitive and affective phenomenology that is incompatible with sexual activity and develop a strong preference for it. If Paul's argument is correct, cognitive modeling of the subjectivity of a monk-future cannot do what you think it can. You simply allow one of your higher-order preferences (the one not to change your first-order sexual preferences) to legislate against that which you do not yet know how to evaluate (monkish preferences). And this runs afoul of the constraint that rational preferences should range over options we are justified in evaluating.

Perhaps sexual, and similarly strong preferences, are special cases where rationality may permit a principle of conservatism for current preferences. Giving up sex

may be closer to giving up food than, say, giving up a certain percentage of sleep or a certain amount of solitude. Just as it is rationally permissible for anyone to prefer eating food to not, it should be rationally permissible for anyone to prefer their current sexual preferences to drastically altered ones (except, perhaps, if they clash with morality). We certainly do not want Paul's inaccessibility argument to prove too much. Paul doesn't think our preferences change so much that we cannot rationally prefer parenting to a lifetime of excruciating physical torture. Some things about our preferences and values stay fixed across personal transformations.

Even if we deem it acceptable, rational deference to certain kinds of preferences can only restrict domains of discovery options so much. Discovery values plus plausible conservation principles are not going to decide between parenting versus foster parenting, for instance, though there are many possible differences in the experiential values to be discovered. There is a general lesson in the vicinity as well. If the goal is to use non-discovery values to pare the discovery space down to one or a few discovery options, what makes us think that we can or should stop the process there? Why not think that we can find some principle or reason to make the decision without having to appeal to discovery values at all?

While I cannot prove that it couldn't happen, it seems highly improbable that my current preference set plus constraints on rationality determine a decision problem where all but one discovery path has been excluded and all that remains is whether to discover X or avoid discovering X. Rather, the more we use non-discovery values to rule out possible transformations as live choices, the more we should raise our confidence that transformational decisions can be made without appeal to discovery values at all.

And, if we are doing this while respecting Paul's accessibility constraint on values, the more we are undermining the problem that motivated the discovery solution in the first place. That is, we are appealing to values other than the experiential (and inaccessible) ones dictated by the ideals of personal fulfillment and self-realization. If other relevant values besides experiential or discovery ones are ready to hand, then it is hard to see what motivates Paul's problem in the first place.

Augmentation Strategy #2 Another augmenting option would recommend values or preferences more germane to discovery itself. These would be rules for 'virtuous' discovery, so to speak, rather than elimination of discovery options via non-discovery preferences. One example of such a rule might be to prefer the kinds of discoveries that people around you choose: the culturally salient options for discovery. Perhaps becoming a celibate monk just isn't a culturally salient route to transformative discovery. Cultural salience correlates with a lot of desirable features: camaraderie, support, sharing. Humans are a social species that do things in social contexts. Discovery should be thought of in a social way. We value society with other transformed people. The consideration of discovery value should respect that.

If preferences for fellow travelers in transformation is all that matters, solitary endeavors such as being a hermit will be ruled out, but this will be a small percentage of the total transformation space. If the preference is for continued society with one's own culture or sub-culture or family or friend-group, it is hard to see how these are really virtues properly thought of as attaching to the value of discovery itself. The person who prefers to travel to foreign lands and pursue transformation among a new culture cannot be violating rules of virtuous discovery. Once again, too, the preference for discovery-in-

one's-own-cultural-sphere is not likely to significantly delimit transformational possibilities. Career choice remains a good example here. The cultures in which personal fulfillment and self-realization are such high ideals are also ones where, often, many differentially transformative career options are available.

Perhaps the most plausible candidate for a rule of virtuous discovery is that discovery choices not be self-undermining. I suppose there are multiple possible ways to be self-undermining. With many values, one obvious way for it to happen is when the acquisition of some prevents the acquisition of more. For many values, more is better (up to certain limits, often). Pleasure is a good example. Other things equal, x+1 units of pleasure is better than x units. But if one way of obtaining pleasure prevented you from gaining more pleasure, or diminished your capacity to do so, that would be a "bad" way of obtaining pleasure--it would be in some sense contrary to the aims of obtaining pleasure. This is one reason certain types of drug use are risky or dangerous. They are pursuits of pleasure that can ultimately destroy or diminish your ability to obtain pleasure.

It is plausible that the value of discovery is like this. We are not satisfied with one discovery. We are usually not satisfied with one, or a few, revelatory discoveries. If, on the basis of past experience, we value the discovery involved in personal transformation, it seems we must value more than one such discovery. If so, then we should choose to discover personal transformations that do not inhibit our ability to discover further personal transformations. We should choose against personal transformations that require taking on moral obligations and commitments that would be violated by another personal transformation. From the point of view of transformation discovery, we should

choose against transformations that are likely to extinguish or reverse the preference for discovery of transformation. But I take it that the value of discovery was not proposed to be a means for ruling transformative choices out in this kind of way.

Many personal transformations, e.g., parenthood, require long-term, even life-long, commitments and responsibilities that rule out further personal transformation. So the value of discovery might actually end up recommending against certain transformations that Paul thought it could rationally justify. Becoming a parent may be one. To rationally choose transformations requiring such commitment, we would need to appeal to values besides discovery. Unless, of course, there is a brute preference for discovering transformations requiring long term commitments. I consider appeals to moves like that next.

Digging-In What about an approach that 'digs-in' on basic discriminating discovery values? My preference for discovery of X over Y is basic. And it doesn't need further support. After all, rational decision theory must take preferences as basic at some point. Furthermore, we do accept some basic discovery preferences. Some people just prefer to discover truths of social insect ecology to discovering truths of black hole formation. Some people prefer to discover the experience of skiing to the experience of playing basketball. That's all there is to it. Perhaps my argument that discovery is silent threatens to prove too much. It threatens to undermine all the virtues of discovery that were previously noted. I can't claim both the earlier noted virtues and that it is impotent to guide us.

Some desires and preferences may indeed be basic or 'brute'. A drive to find out about the world may indeed be one of them. But such a 'discovery drive' is something

different than a preference to discover a specific kind of personal transformation over another. Discovery can get us into trouble and so the desire to do it needs guidance. The kinds of cases mentioned in the previous paragraph are not usually best explained by a basic and unquestionable preference for the discovery of X over the discovery of Y. Rather, people's personal histories, including past experiences, will play a role in explaining what shaped such preferences. To the extent that basic desires play an explanatory role, they are more likely to be ones that are not best described as such fine-grained discovery preferences. Given our biology, in most cases, a simple desire to parent will be far more comparatively basic than a desire to discover parental transformation. The claim that there are basic yet rational preferences for discovery of some kinds of personal transformation to others just isn't plausible.

The fact that discovery cannot guide us in realistic transformative decision situations is reason enough to continue searching for other values that can. But it faces another problem that is worth considering because meeting it may be a significant challenge for any candidate value-bases for transformative decision making. I turn to that problem, the "gravity problem", next.

Gravity[51]

Prior to hearing Paul's arguments for the unavailability of a whole class of values, one might be taken aback by the suggestion that we should base the parenting decision on the value of discovery. There is so much at stake: hundreds of thousands of dollars; thousands of hours of sleep; opportunities of all kinds; the existence of human beings. Are we really to believe that all of it could turn only on what we think about learning

51 I am grateful to Bernhard Nickel for discussion of ideas that led to my framing of this problem.

new things? It can seem as though the recommendation does not respect the gravity of the decision: almost like flipping a coin to decide a Supreme Court case.

Remember that the experiential values to be discovered remain unknown prior to revelatory experience of them. But they retain their potential "heft" in decision making, as when we were thinking about the problem before the consideration of discovery values. This means the discoverer will still have to live with the discovered experiential values and the person who rejects transformation must live with rejecting the values even if she does it on the basis of a preference against discovery.[52] The whole reason that personal transformation is supposed to be a problem that does not arise for the discovery of new tastes is that the transformation has a kind of weightiness and level of consequence that the other does not. The weightiness concerns the very shape and contour of one's personal identity. Can the value of sheer discovery, even discovery of personal transformation, really rise to that level of seriousness?

To make this point more precise, let's compare how the discovery solution fares with novel kinds of tastes, like durian. As Paul herself notes, we have tried new fruits before and that gives us important information (2014, 38-39). We know that the probability of a new taste having a dramatically negative effect on us is very small. We know that any transformation will almost assuredly be very restricted, leaving our core values and preferences intact. Another way of putting this is that we are confident that the inaccessible experiential values of tasting a new fruit are very few in number, and

52 There are several places where Paul writes of the potential for one set of values to "swamp" another set (e.g., 2014, 42-43; 73n.28; 75; 163). This seems to be her version of the issues I am grouping under what I call the gravity problem, though there may be differences. In her terms, then, I argue that it is implausible that discovery values could swamp inaccessible experiential values. At least in many cases, such as parenting.

each one occupies a narrow range of possible values with upper and lower bounds that are themselves small in comparison to our cherished values. Let me explain in more detail.

In the case of durian there are at least two revelatory aspects: the intrinsic character of the flavor, and its place on a distastefulness-deliciousness intensity scale (2014, 35). Let's allow for rich phenomenology and throw in a couple more values that are only accessed once we taste durian for the first time.[53] So we'll suppose there are 4 such inaccessible values. By comparison, the other side of a personally transformative experience conceals far more new experiences. Let's stick with the parenting example. The experience of parenting will include new feelings, moods, and affective states, both positively and negatively valenced. If Paul is right, it will contain all kinds of new cognitive phenomenology--new kinds of thoughts and understanding of the world, relationships, and so on. There will be new desires to deal with. Perceptual experience may change as well. You may begin to see things as threats to your children, for instance. The new kinds of experiences Paul thinks we may have are typed very finely. Cognitive concern for children is revelatory in comparison to cognitive concern for loved ones besides children (Paul, personal communication). All considered, I think 30 inaccessible values is probably a conservative estimate of what awaits on the other side of many personal transformations. And that is a sufficient number to make my comparative point.

Now let's think about a value scale. Let it have a true zero point and let me propose some orienting values to provide some kind of anchoring. I'll consider breaking

53 Paul might say there are even more (see her 2014,163). My purpose is to make a comparative point, though, and so this is sufficient as a rough estimate.

one's leg as a really negative event and attending a beautiful, joyful wedding full of friends and family as a highly positive event. I don't claim any but the most rough ball park figures for all quantitative values below. My aim will be satisfied if I can provide a sense for what is involved in weighing discovery values against what could transpire through a transformation like becoming a parent.

Breaking Leg: -400 Wedding: +600

In comparison, I think the possible values associated with the taste of a fruit would be much smaller. In my case, I would say the upper bound would be 10 and the lower bound -10. Given my allowance for four different values, that's a possibility of having an experience of -40 upon tasting durian. In comparison to the point I set for breaking a leg, I think I'm being generous to the power of durian to displease me. In any case, I would need to value the discovery of the experience at something greater than 40 to guarantee that the experience would be worth it for me. That also seems like a lot, but let's grant it for now. So let us assume the following figures for tasting something like durian for the first time:

New Taste: 4 inaccessible values. Lower bound: -10 Upper Bound: +10.

Discovery value for each inaccessible state: +11

Under these conditions, it would always be rational for me to try durian. Now let's try an analogous exercise for personally transformed experience. It is not clear how allowable Paul thinks this is, or what the rules are for doing it (see 2014, 156-7). But it seems to me like there ought to be some way of quantitatively portraying the asymmetry Paul thinks there is between purely epistemic transformation and personal transformation. The numbers I choose are only meant to do that and should not be

taken to imply any unwarranted precision beyond that. Recalling the number of values I conservatively estimated above, and reasoning that the extended experience of parenting may have values significantly greater than those for events like breaking a leg or attending a joyous wedding, I offer the following figures:

Personally Transformed Experience: 30 inaccessible values. Lower bound (for some): -100,000 Upper Bound (for some): +100,000[54]

On this proposal, the experiential values associated with what it is like to be a parent could possibly be thousands of times more negative or positive than strongly valued events (or more! I am happy to defer to experts, actual parents, on this). Now recall that discovery values are supposed to be accessible to decision-making. If we are able to rationally decide on the basis of discovery values, we are currently able to assign them. I confess that I cannot honestly assign values to the discovery of parenthood to compete with the range of possible figures above. But I also find it hard to believe that anyone could. If so, does that mean that it is always irrational to decide to become a parent?

More informally, what I am trying to capture above is the fact that things could go the wrong way with something like parenting, and you could be stuck in a miserable situation. The value of discovering a transformed self in a miserable situation seems like it would be small consolation for the person who has to continuously live the misery.[55]

---

54 I am assigning these values from the point of view of the transformed person. This is something that Paul may object to. But her argument for inaccessible values is not strong enough to show that a pre-transformed self cannot understand that their transformed self could be in a miserable situation. That is what I am trying to convey here.

55 This is so even when discovery is continuous, as surely it is in parenting. The miserable transformed person may discover that there are ever new ways of being miserable. At best that only mitigates the misery.

At the end of the day, one still has to be a parent when the value of discovery is outweighed. Perhaps my proposed figures unjustifiably stack the deck against discovery values. If so, more needs to be said than Paul provides about how discovery values can compete with possible transformed values. Discovery does not render irrelevant the need to anticipate how badly things could go with a decision.

In the absence of more information about how discovery values can compete with the possible ranges of inaccessible ones, the strongest response I can think of for the discovery theorist at this point is to say that the gravity problem lies in wait for any proposed solution to our problem. Furthermore, it is the only solution on the table. How else can we deal with inaccessible values other than to ask ourselves how much we value rendering them accessible? We can do no better than to work with the only tool we've got. In the next section, I will propose an alternative. I will argue that it does not fall prey to guidance problems and propose a way in which it may deal with the gravity problem.

Another Way: Choosing the Most Worthwhile

If you want to make a transformative choice, we have so far ruled out doing so on the basis of experiential values, whether cognitively modeled or scientifically determined. I have also argued against doing it on the basis of the discovery of those values. However, even if you have accepted the arguments against these bases for transformational decision making, you may still have a lingering sense that you can make such decisions rationally. You might not be able to shake the feeling that, after all, of course you know that you strongly prefer parenthood to monkhood or to remaining

childless (or visa versa!) and that it would not be irrational to act on that preference. I think there is something right about that feeling and I aim to vindicate it in this section.

Imagine a forced-choice situation where you have to choose between becoming a parent and spending the rest of your life stripping paint. I'm not talking about a painting career, where stripping paint is part of the job. And you get to do things besides your job. I'm talking about a life devoted to stripping paint, with some time set aside for eating, rest, and the other necessities. You may suppose that you can get really good at stripping paint. Perhaps you will even develop new techniques for doing it more quickly, efficiently, or safely. You can be as playful as you want with it as well, to help fend off boredom. You may also suppose that it lends to some larger goal. Perhaps you are ridding buildings of toxic paints or preparing them to be repainted by the world's greatest paint team. Why might you, or anyone, prefer parenting to paint-stripping?

The answer, I propose, is that many people would prefer parenting to paint-stripping, even ideal paint-stripping, because it is believed by them to be inherently more worthwhile.[56] It is because they believe that doing things like comforting one's own child, or helping one's child learn a valuable life lesson, is simply more worthwhile than attaining a new personal best, or even a world record, in the time it takes to strip 500 square feet of paint. Being a good parent is more worthwhile than being an ideal paint-stripper and so is striving to be the former over the latter. People value being

56 People often use 'meaningful' to describe some of the values I am after here. I like the word better than 'worthwhile' but some people think it has more experiential connotations than 'worthwhile'. If uses of 'meaningful' capture any aspects of intrinsic worth of identities or activities that are not captured by 'worthwhile', my use of 'worthwhile' is intended to cover them as well. I will stick with just 'worthwhile' for uniformity and simplicity reasons.

parents over being paint-strippers independently of the experiential values associated with them.[57]

Before commenting further on the consequences of the example, let me make a few remarks on the relevant comparisons. First, devoting one's life to paint-stripping is just as much a personally transformative choice as any other that Paul discusses. A life that revolved almost exclusively around paint-stripping would be very different from one we know and equally hard to imagine, if not more so, as any transformation we actually contemplate undertaking. Since it would involve a personal transformation, we can say all the same things about it. For example, we may suppose that our core-preferences could change in such a way that we come to strongly value the experiential qualities of being a devoted paint-stripper. Our phenomenological sense of self would change, and we would discover much that is novel in expert-paint-stripping-experience. My point is that judgments of worthwhileness justify the rationality of people regarding transformed preferences for such experiences as warped and contorted, not just merely different from their own. Insofar as people could admire the virtuoso paint-stripper, it would be under conditions of forced paint-stripping, where the person is making the best of her situation.

The extremity of the example is only supposed to highlight what I think is a basic fact of human evaluation: that we make judgments of inherent worthwhileness which

---

57 At one point Paul makes reference to "life satisfaction" and "meaningfulness", which may or may not be part of what I am advocating for here. But she does seem to compare them to non-phenomenal factors when she says that such are not under consideration, especially if they do not "swamp" experiential values (2014, 73 n. 29). My project can be understood precisely as trying to meet this "swamping" problem, which I think unavoidable for any potential solution including her own (as my discussion of the gravity problem should show). Additionally, I do not think all "external, non-phenomenal factors", including worthwhileness are moral or ethical factors (see more below). But I do consider the examples she uses (environmental impact and population control) to be ethical factors.

inform our decision making. These sorts of judgments are common place. The skateboarding trick itself is cool; the three-point shot itself is impressive; the piano playing itself is graceful and masterful; climbing the mountain is itself an achievement. We may think of the myriad ways in which human performances can be intrinsically valuable as determinants of the value-determinable I am calling worthwhileness.[58] Crucially, people are often clearly motivated by these values. They are motivated by the potential to achieve gracefulness, masterfulness, finesse, creativity, expressiveness, and so on. Usually, in the pursuit of skills and other valuable activities, people are not inhibited by worries over how the associated phenomenology might turn out.

Now, I'm not saying that people don't, much less shouldn't, worry about the kinds of phenomenology that Paul argues are inaccessible. But it would be equally implausible to hold that the kinds of judgments I've been illustrating are absent from transformative decision making. I do not need to claim that such judgments are fully objective, non-relativistic, or that fully rational choices must agree on what is worthwhile. My point is only that we can and do make such judgments and allow our preferences and choices to be guided by them.

Let me stress that I understand these values of worthwhileness to be non-moral or non-ethical. That is, morality's mandates and prohibitions do not guide in matters of worthwhileness. Morality leaves open the possibility of rational disagreement over what is worthwhile and to what degree. It also allows rational differences of opinion over how much decision making weight to place on values of worthwhileness vis-a-vis other kinds

58 Thanks to Ned Hall for suggesting this way of putting it.

of values, e.g., experiential ones. It is not part of my project to argue that everyone ought to base all major life-decisions solely on judgments of worthwhileness.

Before moving to possible worries, let me say that actions and activities are not the only things we value in this way. We similarly value the practical identities and social roles and statuses we come to inhabit by means of transformation. For some, being a public servant of some kind is a worthy thing to be and to aspire to be. Relationships themselves are also so valued. A parent-child relationship, to many, is to be valued for its own sake. Judgments concerning the inherent worth of identities, actions, and relationships should all be part of the transformative decision making process. In the next sections I will consider potential problems for these 'intrinsic worth' values.

Inaccessibility Redux?

To be a viable alternative to Paul's proposal worthwhileness values still need to meet some basic constraints of rationality. They cannot be inaccessible, for instance. After all, if Paul's interpretation of our current decision making process is correct, we often base our decisions on that which we have no right to. Perhaps we are in the same boat with respect to the worthwhile and meaningful. Are such values also inaccessible?

Values of inherent worthwhileness cannot be in exactly the same boat as Paul's experiential values. They cannot be inaccessible in the same way because they are not values attaching to phenomenal experiences. There may be an analogous problem, however, because, as in our usual example, decision-makers have not yet been parents nor performed parenting actions. How can they make judgments about whether these

things are more worthwhile than being or doing something else? On what basis do they do so?

This is not the setting to attempt a full-fledged epistemology of judgments of worthwhileness. What little I want to say about it here is by way of addressing an important response open to the experiential values theorist. The response insists that there can be no rationally justified judgments of worthwhileness that are not at least partially based in the experience of worthwhileness. I can allow this general principle but deny that it gets the experiential values theorist what she needs. What she needs is that there can be no rationally justified evaluations of the specific kind of worthwhileness attaching to practical identities and actions that are transformative from one's current perspective. I deny this because our powers of generalization and projection are much less restricted with respect to values of worthwhileness than with respect to experiential values. I argue this by way of noting some crucial differences between how we judge worthwhileness and how we judge experiential values.

Difference #1: Recognitional Capacities Consider the recognition of incredible athletic feats and virtuoso musical performances. One need not have the slightest previous experience doing such things oneself to be able to recognize the skill and greatness of such acts when we witness them. Likewise, I claim, one need not have been a parent to recognize the worthwhileness of certain parental acts. Thus, even if internally experiencing the value of worthwhileness as such is necessary for desiring it and forming preferences for it, it doesn't follow that the internal experience of the worthwhileness of X is necessary for being able to recognize and prefer the worthwhileness of X.

Difference #2: Fineness of Grain Experiential values are much more fine-grained than those of worthwhileness. Paul emphasizes the richness and complexity of experiential values. And she claims that the cognitive phenomenology involved in the worry over loved ones prior to having children cannot be adequately projected to know what it is like to worry over one's own children. In contrast, one need only have experiential confirmation that loving familial relationships are worthwhile to prefer lives filled with loving familial relationships to ones without them on the basis of worthwhileness. More strongly, while parental phenomenology and its value is inaccessible to children not themselves parents, the worthwhileness of strong, healthy parent-child relationships is not. A child can justifiably infer, I claim, from the worthwhileness of her side of the relationship to the worthwhileness of the relationship as a whole, and to that of the other side.

Difference #3: 'Theory'-Ladenness In arguing for inaccessibility, Paul usually mentions the immediate reactions we have to revelatory experiences and how we cannot anticipate them. In addition, the descriptions people offer are simply not up to the task of informing us what they are like. It seems as though experiential values are reactively formed upon initial encounter. There is no other way for them to be formed. In contrast, evaluative judgments of worthwhileness are guided by 'theory', understood broadly. We are socialized into cultures and sub-cultures full of traditions and judgments about what is worthwhile and what is not. Anyone who has been prodded by their parents to pursue some extracurricular activities rather than others (e.g., music practice over television), or certain career paths rather than others, knows this well. Our judgments of worthwhileness are informed by ideals we are raised with--one's that people can

critically engage with. Forming preferences about what is worthwhile is not like initially formed taste preferences for chocolate over vanilla.

Some first-personal experience of things that are inherently worthwhile may be necessary for rationally projecting and generalizing judgments of worthwhileness. But it does not follow that one cannot rationally judge the worthwhileness of X, for some X, without first-person experience with X. I think the considerations above present a plausible case that we can and do make some such judgments rationally.

What of Cognitive Modeling?

Paul's focus on experiential values depends on an intuitive feature of decision-making for everyone. Faced with a decision resulting in different and unknown futures, we attempt to get as much information about those futures as we can. One natural, even instinctive, way to do this is to imaginatively and cognitively model the future, using past experience as our guide. Does my alternative position require me to reject that we do this? I surely do not want to deny that we engage in such modeling. But do I deny that we are justified in doing it or that it contributes anything to our decision making? Insofar as cognitive modeling of our potential futures involves modeling the experience of our futures and how we would evaluate them, Paul herself has established a weakness in our usual method of decision making. And, insofar as our cognitive modeling is aimed at accessing future experience, does that demonstrate how little attention we pay to values of worthwhileness? After all, one might think that you don't need to project the future to make judgments about something's worthwhileness.

Of course, I need not deny that personally modeling potential futures is involved in our decision making to hold that there are additional factors to the process. Factors

besides ones based on modeling certainly play into our decisions. But I want to deny that modeling the future is only about, or exclusively for the purpose of, accessing potential phenomenology. When we model the future, among that which we model are objective events. Besides the inherent implausibility of denying this, Paul needs it for her own characterization of what we do when we model the future. She thinks the experiences we value, or attempt to value, are of objective events, and stipulates that the experiences she is interested in must be largely veridical, or reflective of actual states of affairs (2014, 11-12). Presumably she would also say that when we attempt to model our future phenomenology, we do so in response to what we expect will happen.

But imagining objective events does not just call to mind the experience of those events. It can also call to mind how much we value those events. And this is precisely what I want to claim is one of the things that some people do when imagining transformed futures. People imagine the actual actions they will perform in alternative futures and ask themselves which ones are inherently more worthwhile. I can imagine parenting activities and ask myself how much I value their worthwhileness and whether I want that worthwhileness to be a part of my life. There are other important things we are trying to do when cognitively modeling alternative futures. But this should be enough to deflect any idea that the fact that we use cognitive modeling to guide decision making is a threat to the claim that considerations of worthwhileness are also engaged.

Gravity and Guided Transformation

With regard to the problems I raised for discovery values, I think it should be evident now that the guidance one does not really arise for the kinds of inherent worth values I've been advocating. The motivating examples I've appealed to illustrate well

how such judgments usually aim at comparing and discriminating the values of different things. Indeed most systems of evaluation aim at doing just that. It is a peculiar feature of the discovery value that it cannot do that. But all proposals for dealing with how to weigh or balance out values on the other side of transformation must seriously grapple with the gravity problem. What sorts of values that we currently have access to could possibly help us think about how to deal with inaccessible values that we would have to live with if we undergo transformation? I offer a speculative proposal for how values of worthwhileness can rise to the occasion. It is based on a feature of the transformative process that Paul does not explicitly deny, but also doesn't seem to take seriously in her work thus far. My proposal is based on the idea that the transformative process can be thought of more on a model of training oneself to acquire new tastes than on a model of accepting one's initial reactions to new tastes. We may commit to a project of self-conditioning to value the experiences that accompany what we judge to be most worthwhile to do. Before getting to the details, though, let's recall the nature of the gravity problem.

For the discovery value theorist, I raised the complaint that you still have to get on with the business and responsibilities of parenting once you've discovered what it is like. If it turns out to be pretty miserable, the thought that: "Well, at least I can say that I can keep discovering what it is like to be transformed by parenting" seems like inadequate solace. Can the value of worthwhileness offer better consolation to the miserable parent? This is the issue my self-conditioning strategy is designed to address.

The idea of discovery strongly suggests receptivity and lack of control. One discovers what is already there, independently of himself. When we set out to discover

something, we want to find out about X, not dictate concerning X. The idea of revelatory experience conforms to this model. We want to find out what durian tastes like and how we will respond to it: how delicious (or not) we will find it. But these features of revelatory experience are quite misleading if used as a complete model for what transformation over time, epistemic or personal, is like. This can be seen by considering the phenomenon of acquired tastes which I turn to next.

Revelation and Acquisition

Revelatory experiences are not necessarily very informative or psychologically determinative. Take the initial taste of wine. There is a sense in which my first taste of wine was revelatory (and so epistemically transformative). I can attest that it was new and, in many ways (though not all), unlike anything I had ever tasted. It was far more different from grape juice than I had expected. Still, my initial experience was under-informative concerning the taste of wine. Today I know much more about what wine tastes like than I did after my first few glasses. Tasting wine is a much richer and complex experience for me now than when I first drank it. My epistemic transformation due to my initial encounter with wine was real but, in many respects, really negligible. To extend the revelation metaphor, it was akin to having a few of the middle books of the Bible revealed. Further revelation would bring both additional information and more sense-conferring context. Epistemic transformation increased as my experience with wine increased. Yet, I don't recall a revelatory moment when wine became rich and complex for me. I only recall my initial encounter with it. Revelatory experiences may be the most memorable because of the way our psychology works. But, for the same reason, they may also be much less important to understanding how psychological

transformation works. This is even more apparent when it comes to evaluation of transformed experience.

The value I attached to my revelatory encounter with wine was also not a good guide to how I would value my later, transformed experience with wine. I greatly disliked it the first time I tried it. I enjoy it now. But not enough to put in the effort to be able to distinguish all but a few kinds. By contrast, I've developed a more 'refined' taste for beer and its different varieties, even though I disliked it more than wine the first few times I tried it. The psychological transformations I underwent with respect to wine and beer were not predictable for me based on the revelatory experiences I had with them. Many people report similar experiences. For them, alcoholic beverages are an acquired taste. For whatever reason, they were determined to come to enjoy them. And they were successful in bringing about this taste-transformation in themselves.

Acquired tastes are examples of psychological transformation (both cognitive and conative), which are not well understood on the model of initial reactions to epistemic revelation. Nor is the subject's relation to their own transformation best described merely in terms of discovery. Rather, psychological subjects take on commitments to come to value experiences in a certain way, training themselves to do so. They help guide the formation of their preferences, and actively learn more than revelatory experience teaches. Culinary tastes are not the only things people learn to like, of course. They learn to like all kinds of music, art, activities, ideas, people, and on and on, that would not be predicted from revelatory encounters. Can our ability to actively train our preferences for such things serve as a model for deep personal transformation? Can we commit to actively becoming the kinds of people that value and enjoy transformed lives

with rational expectation of success? If so, we could resist the idea that the experiential values of a transformed life simply await our discovery. Instead, we may adopt the attitude that, whatever the character of the transformed phenomenology turns out to be, we can actively mold our preferences to find them sufficiently agreeable.

Of course, we don't undertake wearying projects of re-training our core preferences for the heck of it. We need motivating ideals to do so. My suggestion is that ideals to be and do the most worthwhile things can fill such a motivational role. If we determine that, say, becoming a parent is the most worthwhile way to go, we may commit ourselves to self-conditioning for the sake of acquiring the ability to value the experience of being a parent and doing parental things with sufficient satisfaction.

An immediate concern for my proposal is whether we can undertake the kinds of self-conditioning I am suggesting with any kind of rational expectation of success. Certainly, success is not guaranteed and so it is a pressing question how one might assess one's probability of success. That question will have to await another occasion, however. I do not have the space to mount a full defense of my proposal here. Let me suggest, however, that considerations besides likelihood of success may sometimes be more important. For people may rationally want to test and prove the strength of their character or resolve to themselves. Attempting to train yourself to 'acquire a taste' for what you judge most worthwhile seems precisely like the kind of thing worth testing your mettle on, if anything is. Of course, this suggestion also demands further development. I only hope to have made the self-conditioning strategy seem like an option worth thinking more about.

Conclusion

Transformative experience offers much to philosophical reflection, and Paul's work on its challenges to rational decision making shows how fruitful such reflection can be. Her articulation of the problem due to the inaccessibility of values necessary for transformative decision making, and proposed solution, illuminate our decision making situation. If she is right about the problem, as I think she is, it is both important and urgent. I have argued that her solution cannot meet the demands of the problem. In its stead, I have argued for a shift in focus to non-experiential values, ones of inherent worthwhileness, that can motivate active adaptation to experiential values. There are surely problems for this proposal that I have not yet anticipated. But I hope to have made the case that it deserves further investigation.

BIBLIOGRAPHY

Allen, J. 2009. *The Lives of the Brain: Human Evolution and the Organ of Mind*. Harvard University Press.

Aristotle. 1923. *Metaphysics*. Trans. W. D. Ross. London: Methuen & Co.

Atherton, M. 2007. Locke on Essences and Classification. In L. Newman, ed., *The Cambridge Companion to Locke's "Essay Concerning Human Understanding"*. Cambridge University Press.

Atran, S. 1999. The Universal Primacy of Generic Species in Folkbiological Taxonomy: Implications for Human Biological, Cultural, and Scientific Evolution. In Wilson, ed.

Atran, S., et al. 1997. Generic Species and Basic Levels: Essence and Appearance in Folkbiology. *Journal of Ethnobiology*, 17, 22-45.

Banaji, M., and Greenwald, A. 2013. *Blindspot: Hidden Biases of Good People*. Delacorte Press.

Barnes, E. Forthcoming. What You Can Expect When You Don't Want To Be Expecting

Bayne, T. and Montague, M., eds. 2011. *Cognitive Phenomenology* Oxford: Oxford University Press.

Boulter, S. J. 2012. Can Evolutionary Biology do Without Aristotelian Essentialism? *Royal Institute of Philosophy Supplement*, 70, 83-103.

Boyd, R. 1999a. Homeostasis, Species, and Higher Taxa. In Wilson, R. ed., 141-186.

Boyd, R. 1999b. Kinds, Complexity, and Multiple Realization: Comments on Millikan's "Historical Kinds and the Special Sciences". *Philosophical Studies*, 95, 67-98.

Brown, D.E. 1991. *Human Universals*. New York: McGraw-Hill.

Buller, D. J. 2005. *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. The MIT Press.

Carey, S. 2009. *The Origin of Concepts*. Oxford University Press.

Charles, D. 2003. *Aristotle on Meaning and Essence*. Oxford University Press.

Degler, C. 1991. *In Search of Human Nature: The Decline and Revival of Darwinism in American Social Thought*. Oxford University Press.

Devitt, M. 2008. Resurrecting Biological Essentialism. *Philosophy of Science*, 75, 344-382.

de Queiroz, K. 1999. The General Lineage Concept of Species and the Defining Properties of the Species Taxon. In Wilson, R., ed, 1999.

Dougherty, T., Horowitz, S., and Sliwa, P. Forthcoming. Expecting the Unexpected.

Dumsday, T. 2010. Natural Kinds and the Problem of Complex Essences. *Australasian Journal of Philosophy*, 88:4, 619-634.

Ereshefsky, M. 2010a. What's Wrong With The New Biological Essentialism. *Philosophy of Science*, 77, 674-685.

Ereshefsky, Marc, "Species", *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2010/entries/species/>.

Ereshefsky, M. 2001. *The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy*. Cambridge University Press.

Ereshefsky, M., and Matthen, M. 2005. Taxonomy, Polymorphism, and History: An Introduction to Population Structure Theory. *Philosophy of Science*, 72, 1-21.

Fine, K. 1993. Essence and Modality. In J. Tomberlin (ed.), *Philosophical Perspectives,* Atascadero, CA: Ridgeview, 1-16.

Gelman, S. 2003. *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford: Oxford University Press.

Ghiselin, M.T. 1997. *Metaphysics and the Origin of Species*. State University of New York Press.

Godfrey-Smith, P. 2014. *Philosophy of Biology*. Princeton University Press.

Godfrey-Smith, P. 2009. *Darwinian Populations and Natural Selection*. Oxford University Press.

Gorman, M. 2005. The Essential and the Accidental. *Ratio*, 18, 276-289.

Griffiths, P. 1999. Squaring the Circle: Natural Kinds with Historical Essences. In Wilson, R. ed., 209-228.

Hauser, M. 2001. *Wild Minds: What Animals Really Think*. Henry Holt and Company.

Hey, J. 2011. Regarding the Confusion Between the Population Concept and Mayr's "Population Thinking". *Quarterly Review of Biology*, 86:4, 253-64.

Heyes, C. 2012a. Grist and Mills: On the Cultural Origins of Cultural Learning. *Philosophical Transactions of the Royal Society B*, 367, 2181-2191.

Heyes, C. 2012b. Simple Minds: A Qualified Defence of Associative Learning. *Philosophical Transacttions of the Royal Society B*, 367, 2695-2703.

Heyes, C. 2012c. What's Social about Social Learning? *Journal of Comparative Psychology*, 126, 193-202.

Hull, D. 1986. On Human Nature. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1986, Volume Two: Symposia and Invited Papers, 3-13.

Hull, D. 1978. A Matter of Individuality. *Philosophy of Science*, 45, 335-60.

Hull, D.L. 1965. The Effect of Essentialism on Taxonomy: Two Thousand Years of Stasis. *The British Journal for the Philosophy of Science*, 15, 314-326 & 16, 1-18.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus, and Giroux.

Kant, I. 1798. *Anthropology from a Pragmatic Point of View*. R. Louden, trans., In G. Zoller and R. Louden, eds., *Anthropology, History, and Education*, 2007, Cambridge University Press.

Kant, I. 1786. Conjectural Beginning of Human History. A. Wood, trans.,In G. Zoller and R. Louden, eds., *Anthropology, History, and Education*, 2007, Cambridge University Press.

Kant, I. 1790. Kritik der Urteilskraft.

Kripke, S. 1980/1972. *Naming and Necessity*. Harvard University Press.

LaPorte, J. 2004. *Natural Kinds and Conceptual Change*. Cambridge: Cambridge University Press.

Lewens, T. 2012a. Species, Essence, and Explanation. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 751-757.

Lewens, T. 2012b. Pheneticism Reconsidered. *Biology and Philosophy*, 27, 159-77.

Lewens, T. 2012c. Human Nature: The Very Idea. *Philosophy & Technology*, 25, 459-474.

Lewis, D. 1986. Causal Explanation. *Philosophical Papers ii*, 214-240. Oxford University Press.

Machery, E. 2012. Reconceptualizing Human Nature: Response to Lewens. *Philosophy & Technology*, 25, 475-8.

Machery, E. 2008. A Plea for Human Nature. *Philosophical Psychology*, 21:3, 321-329.

Machery, E. and Barrett, H.C. 2006. Essay Review: Debunking *Adapting Minds*. *Philosophy of Science*, 73, 232-246.

Margolis, E. and Laurence, S. 2013. In Defense of Nativism. *Philosophical Studies*, 165:2, 693-718.

Marx, K. 1846. The German Ideology. Reprinted in R. Tucker, ed. *The Marx-Engels Reader*, 2nd ed., New York: W.W. Norton & Company.

Marx, K. 1845. Theses on Feuerbach. Reprinted in R. Tucker, ed. *The Marx-Engels Reader*, 2nd ed., New York: W.W. Norton & Company.

Matthen, M. 2013. Millikan's Historical Kinds. In Ryder, D., Kingsbury, J., and Williford, K., eds., *Millikan and Her Critics*, Malden, MA: Wiley-Blackwell.

Matthen, M. 2009. Chickens, Eggs, and Speciation. *Nous*, 43, 94-115.

Mayr, E. 1975/1959. Typological versus Population Thinking. Excerpted from Evolution and the Diversity of Life, Harvard University Press. Included in Sober, E. (ed), 2006, *Conceptual Issues in Evolutionary Biology*, 3rd ed., MIT Press, 325-28.

Nanay, B. 2011. Three Ways of Resisting Essentialism About Natural Kinds. In Campbell, J.K. and Slater, M.H., *Carving Nature at its Joints*, 175-97, MIT Press.

Nicholson, D. 2014. The Return of the Organism as a Fundamental Explanatory Concept in Biology. *Philosophy Compass*, 9, 347-59.

Nowell, A. 2010. Defining Behavioral Modernity in the Context of Neandertal and Anatomically Modern Human Populations. *Annual Review of Anthropology*, 39, 437-52.

Oderberg, D. 2008. *Real Essentialism*. Routledge.

Okasha, S. 2002. Darwinian Metaphysics: Species and the Question of Essentialism. *Synthese*, 131: 191-213.

Paul, L. A. Forthcoming. Transformative Experience: Precis and Replies.

Paul, L. A. 2015. What You Can't Expect When You're Expecting. *Res Philosophica*, 92:2, 1-23.

Paul, L. A. 2014. *Transformative Experience*. Oxford: Oxford University Press.

Pedroso, M. 2012. Essentialism, History, and Biological Taxa. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 182-90.

Pettigrew, R. Forthcoming. Transformative Experience and Decision Theory.

Pinker, S. 2003. *The Blank Slate: The Modern Denial of Human Nature*. Penguin Books.

Prinz, J. 2012. *Beyond Human Nature: How Culture and Experience Shape the Human Mind*. New York: W.W. Norton & Company.

Putnam, H. 1975. The Meaning of 'Meaning'. *Minnesota Studies in the Philosophy of Science*. 7:131-193.

Putnam, H. 1970. Is Semantics Possible? Reprinted in *Mind, Language, and Reality, Philosophical Papers, Vol. 2*, 1975, Cambridge University Press.

Rieppel, O. 2010. New Essentialism in Biology. *Philosophy of Science*, 77:5, 662-673.

Robertson, Teresa and Atkins, Philip, "Essential vs. Accidental Properties", *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2013/entries/essential-accidental/>

Samuels, R. 2012. Science and Human Nature. *Royal Institute of Philosophy Supplement*, 70: 1-28.

Samuels, R. 2002. Nativism in Cognitive Science. *Mind and Language*, 17, 233-265.

Sartre, J-P. 1945. *Existentialism is a Humanism*. C. Macomber, trans. Yale University Press.

Sartre, J-P. 1943/1984. *Being and Nothingness*. H. Barnes, trans. New York: Washington Square Press.

Shettleworth, S. 2013. *Fundamentals of Comparative Cognition*. Oxford: Oxford University Press.

Shettleworth, S. 1998. *Cognition, Evolution, and Behavior*. Oxford: Oxford University Press.

Sober, E. 1980. Evolution, Population Thinking, and Essentialism. *Philosophy of Science*, 47, 350-383.

Solomon, R. 1972. *From Rationalism to Existentialism: The Existentialists and their Nineteenth-century Backgrounds*. Rowman & Littlefield Publishers.

Sterelny, K. and Griffiths, P. 1999. *Sex and Death: An Introduction to Philosophy of Biology*. University of Chicago Press.

Strevens, M. 2000. The Essentialist Aspect of Naive Theories. *Cognition*, 74, 149-175.

Sober, E. 1980. Evolution, Population Thinking, and Essentialism. *Philosophy of Science*, 47, 350-383.

Taylor, C. 1985. *Human Agency and Language*. Cambridge: Cambridge University Press.

Thompson, M. 2008. *Life and Action: Elementary Structures of Practice and Practical Thought*. Harvard University Press.

Thompson, M. 2004. Apprehending Human Form. *Royal Institute of Philosophy Supplement*, 54, 47-74.

Thompson, M. 1995. The Representation of Life. in R. Hursthouse, et al (eds.), *Reasons and Virtue*, Oxford University Press.

Tomasello, M. 2001. *The Cultural Origins of Human Cognition*. Harvard University Press.

Wagner, A. 2014. *Arrival of the Fittest: Solving Evolution's Greatest Puzzle*. New York: Current.

Walsh, D. 2006. Evolutionary Essentialism. British Journal for the Philosophy of Science, 57, 425-448.

Weirich, P. 2004. *Realistic Decision Theory: Rules for Nonideal Agents in Nonideal Circumstances*. Oxford: Oxford University Press.

Wilson, R. 1999. Realism, Essence, and Kind: Resuscitating Species Essentialism? In Wilson, R., ed., 187-208.

Wilson, R., ed. 1999. *Species: New Interdisciplinary Essays*. The MIT Press.
Wilson, R., Barker, M., and Brigandt, I. 2007. When Traditional Essentialism Fails: Biological Natural Kinds. *Philosophical Topics*, 35:1&2, 189-215.

Wilson, T. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Harvard University Press.

Wood, A. 2007. Kant and the Problem of Human Nature. In B. Jacobs and P. Kain, eds., *Essays on Kant's Anthropology*. Cambridge University Press.